2021

DISTRIBUTION AND DIVERSITY OF HELIOTHINE AND OTHER LEPIDOPTERAN NUDIVIRUSES

Emrah Ozel
University of Kentucky, emrah@yasambilim.net
Author ORCID Identifier:
https://orcid.org/0000-0001-8278-8767
Digital Object Identifier: https://doi.org/10.13023/etd.2021.449

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Emrah Ozel, Student

Dr. Bruce Webb, Major Professor

Dr. Kenneth Haynes, Director of Graduate Studies

DISTRIBUTION AND DIVERSITY OF HELIOTHINE AND OTHER LEPIDOPTERAN NUDIVIRUSES

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Agriculture, Food and Environment
at the University of Kentucky

By
Emrah Ozel
Lexington, Kentucky
Director: Dr. Bruce Webb, Professor of Entomology
Lexington, Kentucky
2021

https://orcid.org/0000-0001-8278-8767

ABSTRACT OF DISSERTATION

DISTRIBUTION AND DIVERSITY OF HELIOTHINE AND OTHER LEPIDOPTERAN NUDIVIRUSES

Helicoverpa zea nudivirus 2 (HzNV-2) is the only known sterilizing and sexually-transmitted insect virus and causes pathological symptoms in *H. zea* reproductive tissues. HzNV-2 has features that make it a candidate as a *H. zea* (corn earworm) control agent, such as the ability to cause asymptomatic (latent) and symptomatic (lytic) infections and the ability to influence mating behavior of its host to favor virus spread. HzNV pathology has been studied and its genome sequenced, however, its prevalence in natural populations is largely unknown. In this study, we developed and used a low-cost PCR-based molecular survey to investigate HzNV-2 prevalence and found that the virus is circulating in the southeastern United States and reaches up to 40% incidence in some areas. We also discovered a novel nudivirus infecting *Heliothis virescens* populations in some areas of Louisiana and Mississippi. This is the first multi-year study that uses molecular methods and sampling techniques to understand HzNV prevalence in feral Heliothine populations. To further investigate the prevalence of nudiviruses in Lepidoptera, data mining and bioinformatic tools were used to investigate the presence of nudiviruses in NCBI's publicly available sequence databases. This digital survey revealed significant nudivirus prevalence in both *Helicoverpa armigera* and *Helicoverpa zea* populations in Brazil, China, Greece, and Australia. Because the greater genetic complexity of H. armigera nudivirus than HzNV-2, we propose that HzNV-2 originally evolved with *H. armigera* as a host and spread to the Americas as a result of migration and speciation processes that occurred approximately 1.5 Mya. This idea is supported by additional nudivirus detections in a *H. armigera* population and in some *Bombyx mori* cell lines in the old world. Lastly, we sequenced a novel nudivirus that infects *Heliothis virescens* populations and analysis of this sequence revealed a 93.52% similarity to HzNV-2 the genome as well as an incomplete variant derived from the original virus. This in silico finding suggests the presence of a defective interfering particle that replicates and possibly competes with the original Heliothis virescens nudivirus (HvNV). In summary, this work establishes that nudiviruses are present in multiple Heliothine and other lepidopteran species and can be common enough to act as a natural agent controlling these populations. These findings support development of sterilizing nudiviruses as potential agents for novel strategies to control important lepidopteran pests.

KEYWORDS: Helicoverpa zea, Nudivirus, HzNV, Survey, Bioinformatics

Emrah Ozel

12/14/2021

Date

DISTRIBUTION AND DIVERSITY OF HELIOTHINE AND OTHER LEPIDOPTERAN NUDIVIRUSES

By
Emrah Ozel

Dr. Bruce Webb
Director of Dissertation

Dr. Kenneth Haynes
Director of Graduate Studies

12/14/2021
Date

to my family...

ACKNOWLEDGMENTS

I have received a great deal of support and assistance throughout the writing of this dissertation. I would first like to thank my Dissertation Chair, Dr. Bruce Webb, whose expertise was invaluable in formulating the research questions and methodology. I would also like to my Dissertation Committee, and outside reader, respectively; Dr. Ric Bessin, Dr. Jen White, Dr. Mark Farman, Dr. Bruce Webb, and Dr. Arthur Hunt for their valuable guidance throughout my studies.

Finally, I want to thank all researchers who collaborated with and contributed to this study; Dr. Fred Musser, Paul Baker, Dr. Gus Lorenz, Dr. Nick Seiter, Dr. Yves Carreire, Dr. Rob Meagher, Eric LeVeen, Dr. Phillip Roberts, Kelly Estes, Dr. Jon Leuck, Dr. Ricky Foster, Patricia Lucas, Stephen Micinski, Dr. Dominic Reisig, Terry DeVries, Dr. Julie Peterson, Dr. Jerry Goodson, Dr. Randy Boman, Dr. Jeremy K. Greene, Bradley McManus, Dr. Scott Stewart, Dr. Katelyn Kowles, Dr. Lori Spears, Dr. Bryan Jensen, Dr. Dennis Nowaskie, Dr. Philip Walker and Dr. Robert Wright.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Insect Pathogens

Although insects are known to vector microbial diseases to plants and animals, insects are also susceptible to fungal, bacterial and viral pathogens. Some of these insect pathogens such as entomopathogenic nematodes, bacteria, fungi, and viruses have been used as pest control agents against their host insects. The first known insect virus was discovered in European cockchafer (*Melolontha melolontha*) larvae more than 50 years ago (Vago, 1963). After several decades of work, this insect virus was identified as an Entomopoxvirus (EPV) (Miller & Ball, 1998); a close relative to vertebrate Poxviruses. Currently, there are 28 known EPV species classified under Entomopoxvirinae subfamily.

Insect viruses are mainly divided into three groups; *i*) arthropod-borne viruses (arboviruses), *ii*) insect-borne plant viruses, and *iii*) insect-specific viruses (ISVs) (Bolling et al., 2015). Arboviruses are a diverse group of viruses that can replicate in the arthropod host, but also capable of being spread to and infecting a vertebrate host (Bolling et al., 2015, Dader et al., 2017). Many important mosquito-borne diseases including Zika, yellow fever, chikungunya, and dengue fever fall into this category (Achee et al., 2019). Similarly, insect vectors play a key role in infection and transmission of many plant viruses (Dader et al., 2017). On the other hand, insect-specific viruses (ISVs) are restricted to insect cells and are unable to or do not normally replicate in vertebrate cells (Öhlund et al., 2019). ISVs infect both beneficial insects such as honey bees and silkworm and insect pest species, such as bollworm. Among all insect-specific viruses, baculoviruses are one of the most important widespread groups infecting lepidopteran and some hymenopteran pest species.

1.2 Baculoviruses

Baculoviruses are large double-strand DNA viruses that are known to infect more than 400 different arthropod species (Lacey et al., 2001). Among these, baculoviruses infecting forest and agricultural lepidopteran species are relatively well studied in terms of pathology and viral transmission. The Baculoviridae family consist of four distinct nucleopolyhedrovirus (NPV) and granulovirus (GV) genera with phylogenetic data supporting the hypothesis that they evolved from an ancestral DNA virus (Thézé et al., 2011).

Occlusion bodies (OB) are large, proteinaceous capsules synthesized for protecting virions from environmental factors (Hu et al., 1999; Sajjan & Hinchigeri, 2016). Occlusion bodies are comprised of virions embedded within a protein-based matrix that protects the virus particles from degradation in the environment. Most Baculovirus infections produce the OB matrix, which serves as an intermediary for an effective horizontal transmission (Cabodevilla et al., 2011). After being ingested by host larva, the occlusion bodies dissolve due to alkaline environment in the midgut lumen and occlusion-derived virus (ODV) particles fuse with plasma membrane of columnar epithelial cells with virus nucleocapsids entering host cell nuclei where viral replication occurs. In all host groups, except lepidopterans, this primary OB-type infection is localized in midgut columnar epithelial cells or its homologs (Volkman, 1997). Systemic, or secondary, infections occur when budded virions spread to other tissues (e.g fat body, hemocytes, epidermis), predominantly via tracheal networks (Engelhard et al., 1994).

Baculoviruses are large DNA viruses expressing dozens of genes in a coordinated cascade of gene expression. BV genes are grouped as immediate-early, early, late, and very

late genes based on their sequential and coordinated expression (Friesen, 1997). In general terms, immediate-early and early genes prepare the host cell for viral replication by stopping cellular division and suppressing antiviral activity such as induction of apoptosis. The late stage of the BV infection is marked by viral DNA replication and expression of viral structural proteins and assembly of virions that bud from infected cells. Non-occluded budded virus support tissue-to-tissue baculovirus infections within an infected larva. A transition then occurs as cells transition from budded virus production to assembly of OB in the nucleus, which are released into the environment when an insect dies from baculovirus infection and become available to infect other larvae. OB production marks the late stage of virus infection within cells while at the organismal level behavioral changes may occur that promote virus dispersal in the environment. As infected late-instar larvae enter this terminal stage they may climb to the top of plants or edges of leaves (known as tree-top disease) as a result of locomotory hyperactivity caused by viral gene expression (Popham et al., 2016). At this point the larvae can dissolve as chitinase genes break down cuticle to release OBs from infected tissues. OBs remain infectious on plant tissue and in the soil for years until ingested by another larva. This behavioral modification is believed to increase the dispersal of viral OBs (Gasque et al., 2019). This biphasic replication cycle is also referred as primary (OB type) and secondary (budded virus type) infections (Hu et al., 1999).

Infections of baculoviruses can have lethal or sublethal effects. Bombyx mori nucleopolyhedrovirus (BmNPV) is the primary pathogen of silkworm colonies and causes significant losses in sericultural production (Jiang et al., 2021). Many other Baculoviridae members are used against specific pest species. Currently, there are more than 60

Baculoviral insecticides in use worldwide against many costly crop and forest pests, including *Helicoverpa, Autographa, Spodoptera, and Lymantria* species (Beas-Catena et al., 2014). Baculoviruses are also widely used as bioreactors to produce large amounts of engineered proteins and gene products in several baculovirus expression systems (Caron et al., 1990; Elias et al., 2007).

1.3 Non-occluded Baculoviruses

In 1966, a novel virus was isolated from an Indian palm rhinoceros beetle (*Oryctes rhinoceros*) population (A. M. Huger, 1966) and named as Rhabdionvirus oryctes gen. Although the pathology of this new coleopteran virus resembles a Baculovirus infection and rod-shaped virions were produced by this DNA virus, electron micrographs revealed that it lacked polyhedra or a crystalline proteinaceous matrix surrounding the viral particles. Further studies showed that a facultative occluded stage can occur in larval midgut epithelium under some conditions (A. Huger, 1971). Later the virus was placed in the Baculoviridae family and became the type species for Group C, the Non-occluded Baculoviruses (Matthews, 1982) after a complete revision by the International Committee on Taxonomy of Viruses (ICTV).

In subsequent years, other non-occluded baculoviruses were isolated from Coleoptera (K. S. Kim & Kitajima, 1984), Hymenoptera (Bailey et al., 1981), Diptera (Larsson, 1984), Orthoptera (Boucias et al., 1989) hosts, and from other arthropods (Beard et al., 1989; Wongteerasupaya et al., 1995). The Nudivirus genus was proposed in 2006 based on the genomic features of Oryctes rhinoceros virus (OrV) as the type virus (Wang, van Oers, et al., 2007). Later, researchers showed that there is a genetic relatedness between

OrV and Heliothis zea virus 1 (Hz-1) viruses and grouped these viruses under a new genus, Nudivirus, as a baculovirus that lacks polyhedral or OB's. In 2013, Nudiviruses were recognized by the ICTV as a new family, genetically related to but separate from Baculoviruses, based on their similarity to *Oryctes rhinoceros* nudivirus (ICTV Code: 2013.003a-kI). Interestingly, several nudivirus species "facultatively" produce occlusion bodies under some conditions (Bézier et al., 2017).

1.4 Nudiviruses

The origin of the Nudiviridae family traces back to Nudibaculoviridae subfamily (Francki et al. 2012) of Baculoviridae which contained several non-occluded baculoviruses until the subfamily dissolved in 1995 due to conflicting genetic evidence (Bateman & Stentiford, 2017). The Nudiviridae family was proposed in 2007 as a taxon that contains viruses similar to OrNV type species (Wang, van Oers, et al., 2007). In 2013, the Nudiviridae family contained at least 4 members that share 15 or more homologous core genes. The Nudiviridae family consists of 4 genera (alpha-, beta-, delta-, and gammanudivirus) and 11 species. Since 2020, the Nudiviridae family has been classified under the Lefavirales order along with Baculoviridae and Hytrosaviridae families (ICTV Code: 2020.006D). The name of the family derived from Latin word "nudus" which means "naked" and reflects their non-occluded nature. Even though nudiviruses share several Baculovirus core genes, the Nudiviridae family is probably a heterophyletic assembly of non-occluded and facultatively-occluded viruses (Jehle, 2010).

Nudiviruses are large rod-shaped double-strand DNA viruses that can infect various arthropod species. Nudiviruses can be enveloped and/or complexed with nucleocapsid

proteins. With the exception of HzNV-1 and HzNV-2, nudiviruses can also have a tail-like appendage similar to OrNV nucleocapsid appendages (A. M. Huger & Krieg, 1991; Drezen et al., 2012). Genomes of the known nudiviruses contain 98 to 140 open reading frames (ORFs); 33 of those ORFs are shared across the family and 20 of those are homologous to Baculovirus core genes (Wang & Jehle, 2009). One particular gene (*polh/gran*) encodes Baculovirus-like polyhedrin/OB proteins and it is shared among three Nudivirus genera (alpha-, beta-, gammanudivirus).

Nudiviruses and Bracoviruses form a monophyletic clade in large dsDNA viruses group. Bracoviruses are "endogenous domesticated viruses" that are found in parasitic wasps of Braconidae family and composed of two gene clusters, nudiviral genes and proviral segments (Louis et al., 2013). Moreover, bracoviruses are known to integrate into its host genome through a conserved sequence known as Host Integration Motif (Muller et al., 2021). Molecular evidence suggests that Bracoviruses (Family: Polydnaviridae) are evolved from a nudivirus ancestor around 310 Mya as they form a monophyletic clade along with four nudivirus species (Fig. 1.1) (Thézé et al., 2011).

1.5 Heliothine Nudiviruses

In 1978, another novel non-occluded baculovirus, Hz-1, was identified from an established cell line (IMC-Hz-1) (Granados et al., 1978), nearly a decade after the first nudivirus discovery. The IMC-Hz-1 cell line was derived from corn earworm (*Helicoverpa zea*) ovarian tissues and early studies showed that IMC-Hz-1 cell line was nonsusceptible to many insect viruses or showed unexpected declines possibly due to persistent Hz-1 infection (Huger & Krieg, 1991). Further experiments showed that the Hz-1 can

persistently infect a wide spectrum of lepidopteran cell lines causing different levels of cytopathogenic effects (CPE) (McIntosh et al., 2007; Ralston et al., 1981). Moreover, persistent Hz-1 infections can be induced to replicate in presence of some other NPV baculoviruses, especially if these NPVs are UV-inactivated (Kelly et al., 1981). Genomic sequence and features of HzNV-1 were published almost 3 decades after its discovery (Cheng et al., 2002)

1.6 Helicoverpa zea Nudivirus - 2

The first report on HzNV-2 was published in 1995 based on agonadal *H. zea* insects obtained from USDA-ARS laboratory in Stoneville, MS. In that report, several ultrastructural and pathological features were identified and the virus was named as Gonad Specific Virus (GSV) due to its strict localization in gonads and reproductive tissues (Raina & Adams, 1995). Similar to other nudiviruses, HzNV-2 is a rod-shaped, large dsDNA virus lacking occlusion bodies in its reproductive cycle. HzNV-2 replicates in the nuclei of reproductive tissue cells, exhibits either asymptomatic or agonadal pathology, and spreads both horizontally and vertically. The virus also modifies host behavior to improve its horizontal transmission (Burand & Tan, 2006).

1.6.1 Host Biology

Corn earworm (*H. zea* Boddie) is a dispersive, polyphagous and cosmopolitan crop pest that can quickly develop resistance to many widely-used insecticides (Diffenbaugh et al., 2008). Two sister species in the Noctuidae family, *H. armigera* and *H. zea* cause severe economic damage to multiple agricultural crops globally every year. The natural

distribution range of *H. zea* is restricted to North America while *H. armigera* is widely spread in the old world and recently invasive in South America since 2013 (Gonçalves et al., 2019). The distribution pattern of these species did not overlap spatially prior to their recent and ongoing invasion and expansion in South America (Cordeiro et al., 2020) and elsewhere. These population overlaps could generate novel *Helicoverpa* ecotypes that could be more resistant to pesticides as a result of hybridization and genetic introgression between these sister species (Anderson et al., 2018). Moreover, *H. zea* insects can now be found globally as a result of commercial transportation of agricultural products even though it is natural range is restricted to North America. The extend of interspecies hybridization is poorly studied in areas where *Helicoverpa zea* is invasive.

Populations of *H. zea* expand in several ways; short-range (within crop) dispersal, long-range (up to 10 km) expansion, and via migratory movements. *H. zea* moths can migrate several hundred kilometers and move at altitudes up to 2 km (Westbrook et al., 1995) with the help of wind currents. *H. zea* can overwinter in southern states with a hypothetical line around 40° north latitude demarcating the areas where overwintering populations may occur (south) from the areas where immigrant population (north) are re-established every year via migration from the south. Dynamics of *H. zea* populations are usually monitored via light traps or pheromone-baited traps (Hardwick, 1968; Fitt et al., 1989). Under standardized rearing conditions (12h/12h light/dark at 25°C), *H. zea* colonies can produce a new generation approximately in every 30 days. The number of generations per year usually varies between 1 (Canada) to 7 (Florida and South Texas), which also influences the extent of feeding damage. In regions with multiple generations, *H. zea* can infest a different crop each generation based on host availability (Adams et al., 2016).

Corn earworm is an extremely polyphagous insect that feeds on more than 50 host plants including many important agricultural crops, several vegetable plants and weed species. Feeding preference of the corn earworm is determined primarily by host rating and maturity. For larval feeding, corn and lettuce are among the highest quality hosts followed by sorghum, cotton, tomatoes and sunflower (Harding, 1976). Also, females tend to lay eggs on flowering plants which causes ovipositional bias towards mature plants (Johnson et al., 1975).

Corn earworm mating is mediated by sex pheromones that are released by females until copulation occurs. During copulation, a small peptide (pheromonostatic peptide, PSP) is transferred from male to female which inhibits the pheromone release and mate calling behavior (Kingan et al., 1995). Females can lay up to 1500 eggs in the wild and the eggs usually hatch in 2 to 4 days after oviposition. The first instar larva initially feeds on its own egg shell, then grazes on flowering parts of the host then eventually entering the fruit. After its first molt, the larva exhibits cannibalistic behavior that plays an important role in population regulation (Chilcutt, 2006). Last instar larvae stop feeding, drop to the ground and burrow into soil. On average, *H. zea* completes larval development and reaches the pupal stage after 5 (sometimes 6) instars in around 16 days. Pupal development takes 12 to 24 days to complete depending on temperature and soil conditions (Ditman et al., 1940).

The economic damage caused by corn earworm infestation and cost of control reaches several billion dollars per year (Pogue, 2004). Many field studies show that corn earworm infestation can significantly reduce crop yields (Adams et al., 2015, 2016) and the infestations are associated with the spread of *Aspergillus flavus*, a fungus that produces poisonous and carcinogenic aflatoxins (Widstrom et al., 1976). *H. zea* has quickly

developed field-evolved resistance traits to multiple pesticides including Cry toxins. These resistant populations are commonly found across the southeastern United States (Reisig et al., 2018). Recent studies suggest that pesticide susceptibility of *H. zea* diminishes fairly quickly, even rendering pyramided Bt applications ineffective (Carrière et al., 2019). This quick evolutionary response to pesticides is severely limiting management tools leaving only costly options such as promoting susceptibility traits via seed mixtures or block refuges (Brévault et al., 2015).

1.6.2 Structure and Genomic Features

Helicoverpa zea nudivirus is a singly enveloped particle that is roughly 380 nm long and 80 nm wide (Hamm et al., 1996). The HzNV-2 genome is one of the largest among insect viruses containing 231,621 base pairs. The nucleocapsid is a complex rod-shaped structure comprised of nucleocapsid proteins and DNA. A majority of the genes in the HzNV-2 genome are not homologous to any known genes, with 75 of its 113 predicted ORFs annotated as hypothetical genes. The 38 ORFs having homology to known genes encode structural and functional proteins including 16 genes homologous to core baculovirus genes. The HzNV-2 genome closely resembles the HzNV-1 by sharing 93.5% sequence identity (Fig. 1.3) however, 14 predicted ORF regions are missing in HzNV-1 (NCBI Accession: AF451898) and it is 3,532 bp smaller than HzNV-2 genome (NCBI Accession: JN418988). Among these, a large deletion disrupts or deletes ORF90, ORF91 and ORF92 hypothetical coding regions of unknown function. On average HzNV-2 genes are 2,050 bp long based on existing ORF predictions with the total coding sequences constituting 68% of the HzNV-2 genome (Burand et al., 2012). The HzNV-2 genome

contains several genes that are related to DNA replication and repair, transcription, histone synthesis and nucleic acid metabolism as well as 6 tandem repeats, 5 of which are within ORF90 – ORF92 and one within ORF2 (Burand et al., 2012).

Studies with HzNV-1 show that the virus regulates latency via micro RNAs (miRNA) expressed by a persistence-associated gene (*pag-1*) (Chao et al., 1998). miRNA products of this gene inhibit expression of an early-intermediate *hhi-1* transcriptional activator (Wu et al., 2018) that terminates the latent infection and initiates the lytic phase (Wu et al., 2010) and subsequently induces apoptosis of virus infected cells (Wu et al., 2011). These regulatory elements are also present in the HzNV-2 genome and a BLAST search of persistency-associated transcript-1 (PAT1), *pag-1* and *hhi-1* (ORF154 in HzNV-1 genome, Wu et al., 2008) revealed sequence similarities of 95.25%, 100% and 94.87% respectively. This high similarity indicates the presence of similar regulatory elements in the HzNV-2 genome. Besides these regulatory elements, ORF07 of the HzNV-2 genome and ORF145 of the HzNV-1 genome encode a carboxylesterase-type protein and resembles (blastp E-value = $9e^{-27}$) juvenile hormone esterase (JHE) of *Polistes fuscatus* (northern paper wasp).

The HzNV-1 genome lacks several genes that are required for replication in the moth and restrict the virus replication to some lepidopteran cell lines. Establishing cell lines from insect tissues is a complex and poorly understood process that leads to immortalization of some cells that become able to replicate continouosly independent from normal cellular control processes. Some chemical reagents used in the IMC-Hz-1 cell immortalization process, such as sodium hypochlorite (Hink & Ignoffo, 1970) induce mutagenesis in both host and pathogen genomes found in source tissues. Based on this

knowledge, it is likely that HzNV-1 was accidentally created during IMC-Hz-1 cell immortalization process and persisted in the following passages. Moreover, genes deleted during this process are not necessary for viral replication in cell lines.

1.6.3 Pathology

The HzNV-2 virus is sexually transmitted and spreads either horizontally via mating or vertically as an inherited infection from parent to offspring. Oral infectivity is not known to occur naturally even though the virus is somewhat orally infectious in the lab and several per os infectivity factors (pif) and fusion protein genes are present in its genome (Burand et al., 2012). Upon infection, the virus can cause either asymptomatic "latent" or agonadal "lytic" pathology and this biphasic pattern is influenced by the viral titer, with high titer infections favoring lytic infections. In the asymptomatic phase, infected moths do not exhibit pathological signs and reproduce normally, however, latently infected females transmit the virus transovarially. As a result of this progressive viral exposure in the ovaries, progeny from the first- and second-day ovipositions often exhibit asymptomatic pathology while the third- and fourth-day ovipositions usually generate agonadal progeny (Burand & Rallis, 2004).

Lytic infections of HzNV-2 are primarily diagnosed with functional and structural abnormalities in reproductive organs in both females and males. In females, lytic infection causes deformations in gonadal tissue and leads to the formation of a large Y-shaped structure without any visible egg presence (Fig. 1.2 D). Simlarly, males lack seminal vesicles, vasa deferentia and accessory glands in addition to smaller unfused testes. In healthy males, accessory glands produce pheromonostatic peptide, which inhibits the

female pheromone production after mating. These abnormalities cause reproductive sterility (Raina & Adams, 1995). In addition to the internal malformations, lytic HzNV-2 infections usually produce an externally visible "plug" at the females genital opening covered with a dark waxy material (Raina et al., 2000). Further studies revealed that the plug was a mixture of dark-colored granular substance and a light-colored viscous material with high virus concentrations. Moreover, viral replication during pupal and early development prevents formation of normal internal cuticular structures and eventually generating a large Y-shaped and a smaller C-shaped structure (Fig. 1.2, D). The severity of these pathological symptoms are influenced by the infection stage (early instar, late instar or late transovarial) and the degree of sterility varies from partially functional gonads to completely fused non-functional bodies (Fig. 1.2) (Rallis & Burand, 2002a).

Infected males, by contrast, do not exhibit external symptoms and internal copulatory structures develop normal function. However, the testes and accessory glands are either completely missing or significantly underdeveloped. These symptoms indicate that agonadal males can initiate copulation behaviors that support viral spread, however, the infected male cannot transfer spermatophores or auxiliary secretions (Rallis & Burand, 2002b) due to gonadal and reproductive tract atrophy.

1.6.4 Behavioral Modifications

Burand et al. (2005) showed that HzNV-2 infected corn earworm females produce 5 to 7 times more pheromone than healthy females. In lepidopterans, juvenile hormone (JH) titer is closely related to oocyte maturation, calling behavior and pheromone production (McNeil et al., 1995). Also, in noctuids, vitellogenesis and choriogenesis occur

after eclosion and are completely independent from metamorphic events with only juvenile hormone (JH) necessary to initiate gonadotrophic activity (Ramaswamy et al., 1997). On the other hand, juvenile hormone esterases (JHEs) play a central role in regulating the juvenile hormone concentrations by deactivating the JH via a hydrolysis reaction (Khalil et al., 2006). Along with JH, the pheromone biosynthesis activator neuropeptide (PBAN) is another important factor in *H. zea* reproductive system which activates and maintains the pheromone synthesis pathway in pheromone glands tissue (Jurenka et al., 1991; Jurenka & Rafaeli, 2011). Contrarily, several studies showed that the JH actually down-regulates the PBAN responsiveness and thus inhibits the pheromone production (Rafaeli & Bober, 2005; Bober et al., 2010; Choi et al., 2012).

In Heliothine insects, esterases play important roles in insecticide resistance (Achaleke et al., 2009; Li et al., 2013), and reproduction (Gilbert et al., 2000; Khalil et al., 2006). The HzNV-2 genome contains two esterase-like genes, ORF07 (carboxylesterase) and ORF99 (esterase/lipase). The ORF07 sequence resembles the functional motif of insect JHEs and the expression of this gene during lytic phase is likely responsible for JH hydrolysis. Thus, high levels of pheromone production results from PBAN upregulation (Burand et al., 2012). Lastly, as discussed earlier, lytic infections of HzNV-2 cause malformed testes and accessory glands in males rendering the insect incapable of transferring spermatophores and other secretions such as pheromonestatic peptide (PSP). As a result, females receive virions during copulation but exhibit continuous calling behavior which contributes to ongoing attraction of males and mating attempts that spread the virus (Burand & Tan, 2006).

1.7 Project Goals

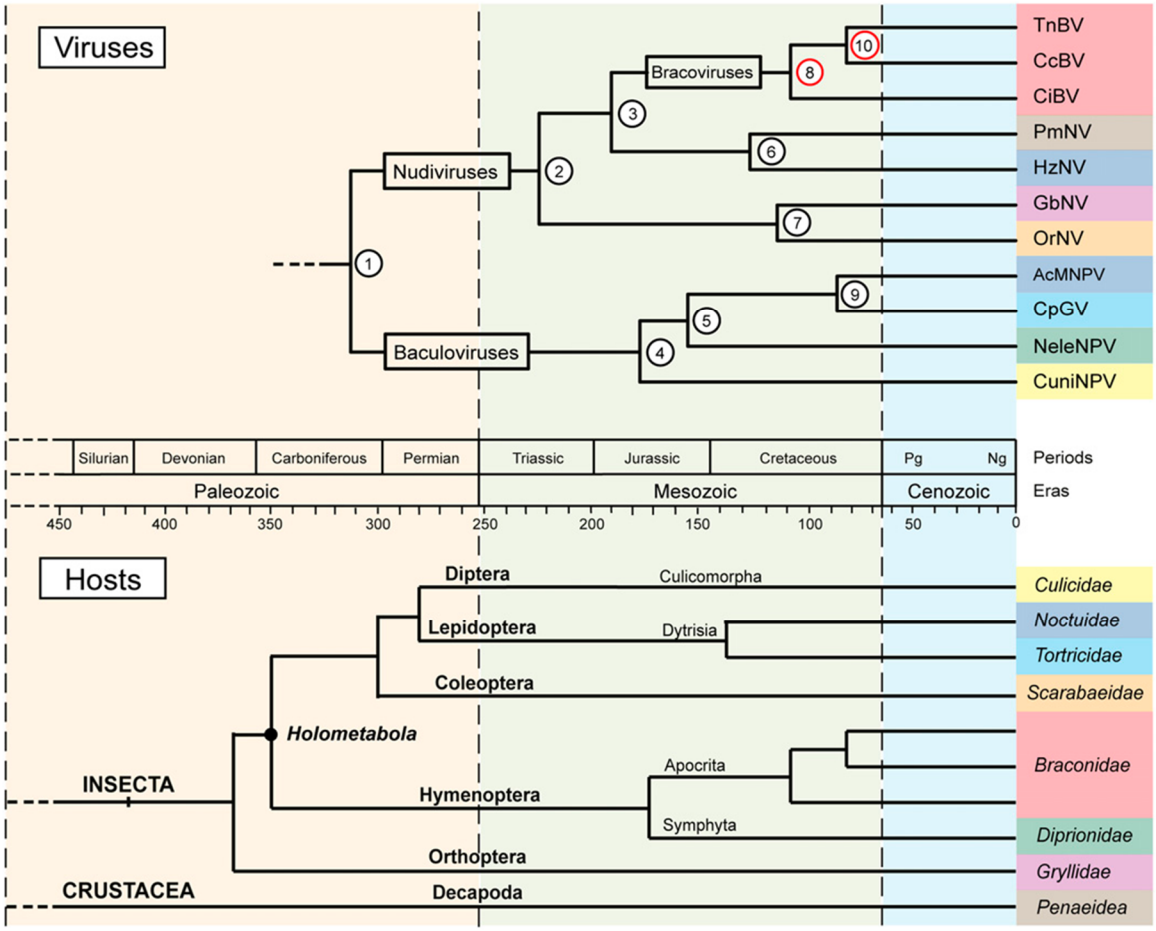
HzNV-2 is the only known sterilizing nudivirus of Lepidoptera and both HzNV-1 and HzNV-2 were discovered by accident based on their effects of infection on insect cells and moths in the lab. There is little published data on HzNV-2 distribution and prevalence in the United States, and there is no information as to whether similar nudiviruses are present in other species. To address this knowledge gap, I surveyed *Helicoverpa zea* and *Heliothis virescens* populations across the US, particularly focused on the United States Cotton Belt because of greater surveillance in this region. This survey was necessary to understand HzNV-2 prevalence in the southest and may contribute to future efforts toward its development and use as a corn earworm management agent. We also were able to survey another cotton pest, *H. virescens,* and detected several nudivirus positives from field-collected samples. These detection results were verified via Sanger sequencing.

Subsequently, I focused on prevalence of nudivirus sequences present in genomic and transcriptomic datasets. To achieve this, I performed a digital survey using next-generation Lepidoptera sequencing datasets deposited to public databases by researchers around the world. This survey produced a number of interesting results, such as strong evidence on *H. amigera* nudivirus in Brazil and evidence that other nudiviruses infecting *Spodoptera* populations and *Bombyx mori* cell lines. I also identified and annotated short nucleotide variations in these datasets for predicting the impacts on transcriptional products. Finally, we screened the entire lepidopteran nucleotide (non-NGS) submissions to more fully characterize and understand sequence homologies and screen for potential viral integration events into lepidopteran host genomic DNA.

In Chapter 4, I have investigated genomic structure of Heliothis virescens nudivirus and compared it with known Heliothine nudivirus sequences. I performed a NGS analysis on one of our strong nudivirus positives that was collected from a *H. virescens* population. This *H. virescens* nudivirus (HvNV) defines a nudivirus that is new to science. It could become the basis for a novel control agent for *H. virescens* management. Lastly, we performed a *de novo* haplotype analysis on this dataset and discovered an incomplete nudivirus strain similar to a defective interfering virus variant that closely resembles a portion of the HvNV genomic sequence. Such variants are known to reduce the efficiency of replication within some baculovirus infections.

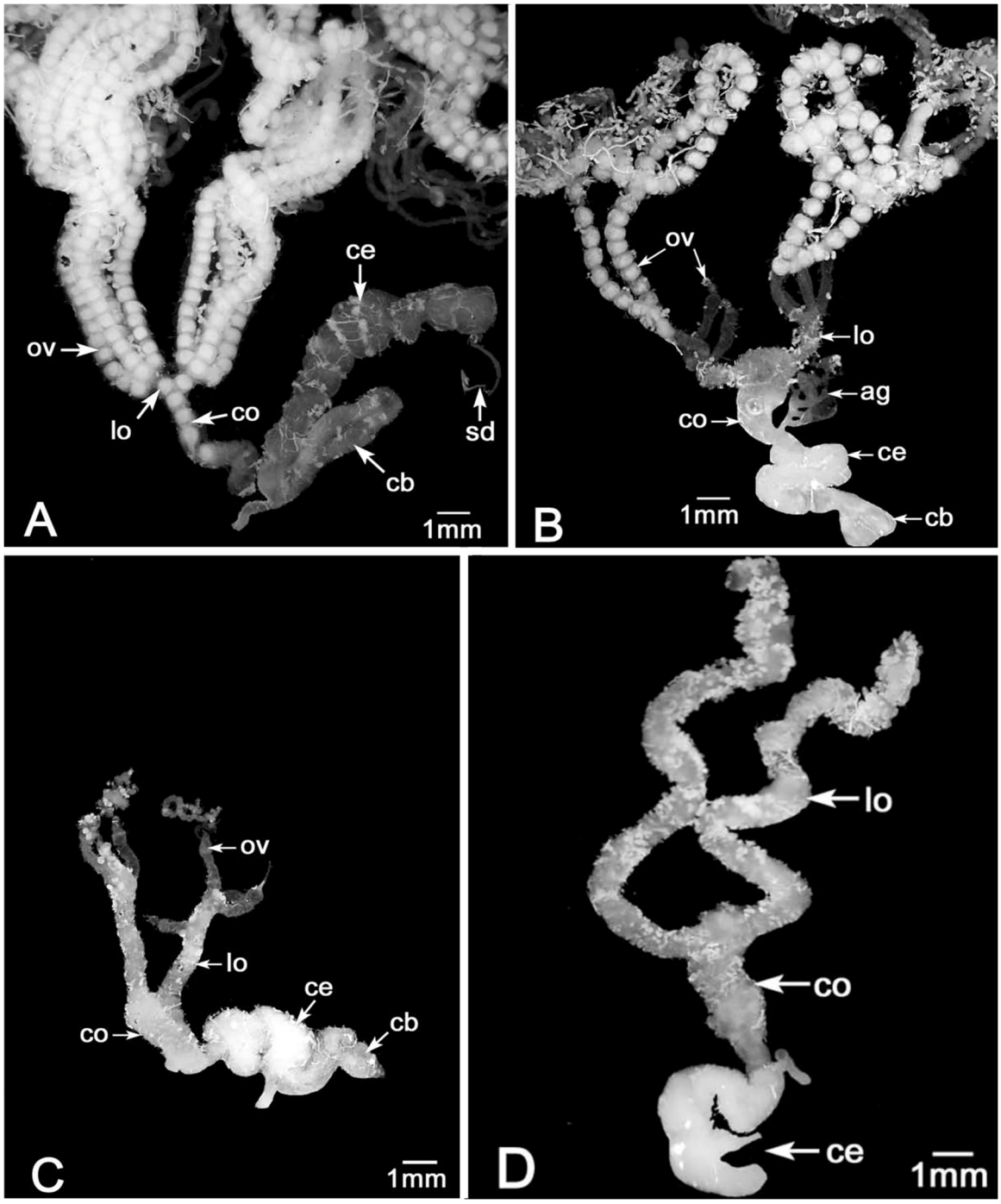
Taken together, this body of work establishes that nudiviruses are globally distributed, vary in frequency in insect populations and are present in multiple genera of Lepidoptera.

Figure 1.1: Evolutionary association of large insect dsDNA viruses (Thézé et al., 2011)



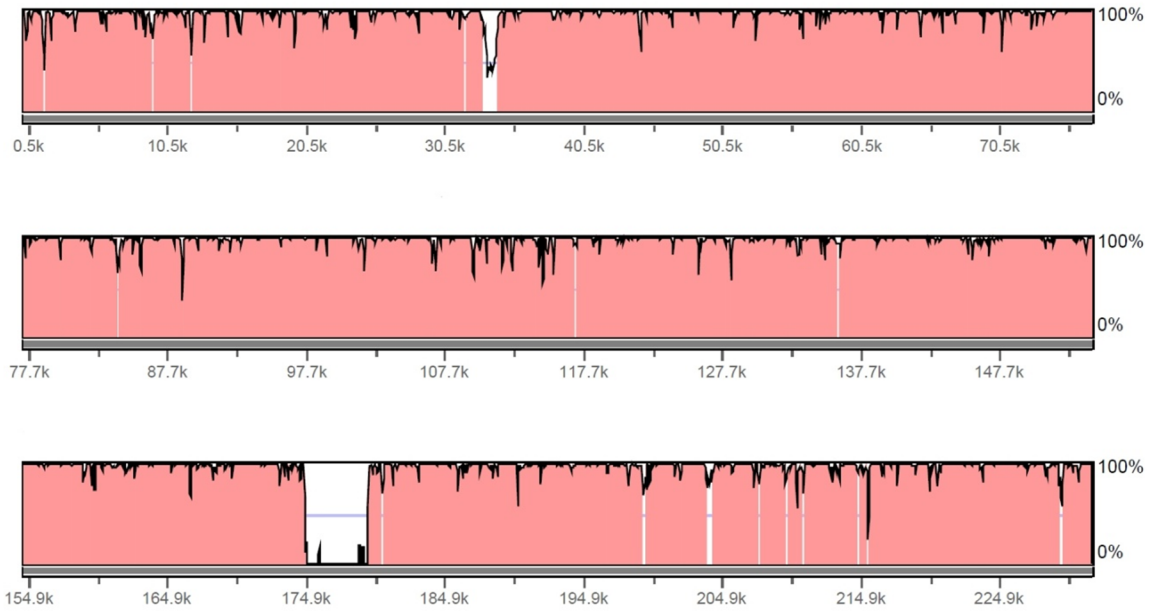
Nudiviruses and Baculoviruses evolved from a common ancestor nearly 300 Mya.

Figure 1.2: Malformation of of adult female *H. zea* reproductive tissues. (A) Healthy adult female, (B) HzNV-2 infected during the late sixth larval instar, (C) during the late fifth larval instar, and (D) transovarially infected adult female. (modified from Rallis & Burand, 2002b)



(Ov: ovarioles, lo: lateral oviduct, co: common oviduct, ce: cervix bursa, cb: corpus bursa, and sd: seminalduct)

Figure 1.3: Whole genome alignment of HzNV-1 and HzNV-2 genome sequence (Burand et al., 2012).



The grey line on the base represents the HzNV-2 genome and the thin black line shows the aligned HzNV-1 genome. Red color denotes a sequence identity greater than 95% and the white sections represent genomic deletions

CHAPTER 2: HELICOVERPA ZEA AND HELIOTHIS VIRESCENS NUDIVIRUS PREVALENCE IN THE UNITED STATES COTTON BELT AND SURROUNDING STATES

2.1 Introduction

Bollworm is a collective term for the Lepidoptera species that feed on and within the cotton boll in the larval stage. This feeding behavior of the bollworm larvae makes it difficult to efficiently manage this pest since the larvae develop contained and protected in the fruit (Moore & Tracy, 2021). The bollworm complex contains major cosmopolitan pest species that cause severe damage to a long list of crop types. Among others, three bollworm species; corn earworm (*Helicoverpa zea, H. zea*), the tobacco budworm (*Heliothis virescens, H. virescens*) and the old world bollworm (*Helicoverpa armigera, H. armigera*) are among the costliest crop pests in the world (Bibb et al., 2018). Larvae of these species also exhibit highly polyphagous feeding behaviors, feeding on more than a hundred different economically important crop and vegetable species including corn, cotton, lettuce, soybean, tomato and tobacco.

Bollworm management becomes more problematic when larvae enter the fruit leaving only few options for control (Tabashnik et al., 2021). Two major microbial bioinsecticides, *Bacillus thuringiensis* (Bt) toxins and baculovirus formulations, have been used to control bollworm infestations. In the United States, the first commercial baculovirus-based insecticide, *H. zea* NPV was released in 1975 by Sandoz (Elcar). Baculoviruses are usually considered as the "natural enemies" of bollworms (Yuan et al., 2021) and they can be engineered to improve their efficiency. One such study elevated the insecticidal activity of a baculovirus strain by incorporating Bt toxins and an insect-specific

neurotoxin in the occlusion bodies (Shim et al., 2013). On the other hand, Bt toxins are popular insecticides of bacterial origin and can be either sprayed (foliar application) or expressed by the plant (transgenic application). Transgenic application is the most effective method for controlling bollworm infestations (Dhillon & Sharma, 2013; Little et al., 2017).

Viruses are important entomopathogens with more than 1100 known strains that infect over 20 insect families (Grzywacz, 2017). Almost half of these insect viruses are members of the Baculoviridae family, which form occlusion bodies to protect the virions from environmental factors (López et al., 2018; Slack & Arif, 2006). Similarly, Nudiviruses infect a diverse group of aquatic and terrestrial arthropods such as Asian tiger shrimp (*Penaeus monodon*), marsh crane fly (*Tipula oleracea*), lesser fruit fly (*Drosophila melanogaster*), and the two-spotted cricket (*Gryllus bimaculatus*). In general, nudiviruses do not form occlusion bodies even though some are known to express polyhedral genes facultatively (Bézier et al., 2017). One well-studied nudivirus, the *Oryctes rhinoceros* nudivirus (OrNV) has been utilized as a pest control agent (Bedford, 2013).

Evidence from comparative genomics indicate that Nudiviruses and Baculoviruses diverged from a common ancestor nearly 310 million years ago (Thézé et al., 2011). These two dsDNA virus families share at least 15 homologous core genes (Burand, et al., 2007) but differ significantly in terms of pathology and life cycle. It is still not clear whether the Nudiviridae family is a monophyletic or polyphyletic as it contains viruses that are capable of replicating in the fat body, midgut and reproductive tissues and even synthesizing polyhedral occlusion bodies under some conditions (Wang & Jehle, 2009; Wang et al., 2012).

Helicoverpa zea nudivirus – 2 (HzNV-2) is a sexually transmitted nudivirus that causes gonad atrophy in corn earworm (*H. zea*) adults. It exhibits a biphasic replication cycle where the infection is asymptomatic during latent phase and has productive, virion producing replication during lytic phase (Burand & Lu, 1997). The productive phase is characterized by a progressive gonad and reproductive tract atrophy internally while usually producing a waxy "plug" at the genital opening on the caudal abdmoninal segments (Burand & Lu, 1997; Rallis & Burand, 2002b). Up to one third of all HzNV-2 infections exhibit lytic (agonadal) symptoms accompanied by reproductive sterility due to physiological and morphological defects in gonads and reproductive tracts (Raina & Adams, 1995). HzNV-2 also modifies host behavior by triggering excessive pheromone synthesis and continuous mating call behavior in females after mating attempts (Burand et al., 2005; Burand & Tan, 2006). The waxy plug formation, which contains high virus concentrations, and changes in moth mating behavior contribute to efficient horizontal spread of the virus in feral *H. zea* populations. Moreover, HzNV-2 is also transmitted vertically through transovarial infection with the phase of the infection (latent or lytic) usually determined by the viral titer (Burand & Rallis, 2004). Despite the factors contributing to the viral transmission, widespread infections are uncommon in feral populations potentially due to an interplay between host immune response, viral dosage and reactivation dynamics.

The first known *H. zea* nudivirus was detected as a persistent baculovirus strain in IMC-Hz-1 cell lines (Ignoffo et al., 1971). Later on, this virus strain was named as IMC-Hz-1 (Granados et al., 1978), then Hz-1V (Huang et al., 1982), and Hz-1 (Burand et al., 1983; Chao et al., 1992; McIntosh et al., 1985) until it was finally classified under

Nudiviridae family (Wang, et al., 2007). More than two decades after the discovery of HzNV-1, another variant, HzNV-2 was detected in a corn earworm colony and initially named Gonad Specific Virus (GSV) (Raina & Adams, 1995). The main difference between HzNV-1 and HzNV-2 is their cellular host range. Whereas HzNV-2 can circulate in wild *H. zea* populations, HzNV-1 replication is restricted to infection of cell lines (Burand et al., 2012). This restriction can be attributed to genomic deletions in HzNV-1 genome relative to the intact HzNV-2 strain. As HzNV-1 was first discovered in an IMC-Hz-1 cell line it is probable that HzNV-1 strain was generated from an HzNV-2-infected gonadal tissue source during the cell immortalization process with subsequent deletions of the virus genome.

Surveying entomopathogens and vector-borne diseases can be time and resource consuming (Shapiro-ilan et al., 2003; Iwashita et al., 2018). Conventionally, FTA® cards (Flinders Technology Associates filter papers) have been used widely as a viable and effective sampling medium for medical and forensic purposes where high quality DNA is needed for subsequent analyses. (Borman et al., 2006; Rogers & Burgoyne, 2000). Despite the advantages of FTA cards, our main focus was to screen wild Heliothine populations for nudivirus presence with no emphasis on downstream analysis so we developed and tested a low cost and effective sampling protocol based on the same filter paper approach.

In this study, we combine several novel approaches to screen nudiviruses in feral Heliothine populations in the U.S. Cotton Belt region and surrounding states. Firstly, we developed an inexpensive sampling method to capture insect DNA required for molecular analysis. Secondly, we established contacts with professionals to collect trap samples using our field sampling kit. The sampling protocol was fairly simple to follow and it allowed us

to achieve a substantial sample size with the help of our collaborators, researchers and field professionals. Unlike a previous study conducted in 8 states in one year (Lupiani et al., 1999), my novel sampling and screening protocol was employed from 2016 and 2020 and allowed us to process over 2000 field-collected samples from more than 50 locations and 4 different host species. Additionally, nudivirus positives were verified with Sanger sequencing. Finally, I believe that the H. virescens nudivirus has potential for managing H. virescens populations and the knowledge produced in this project will support future research on Heliothine nudiviruses.

2.2 Methods

This project was initiated in 2016 to survey nudiviruses in feral Heliothine populations. To achieve this aim, we communicated with researchers and professionals primarily from Southeast and Southwest regions of the United States who monitor Heliothine populations using pheromone traps. I collaborated with more than 30 researchers and field professionals to collecting and ship samples from 39 locations across the Cotton Belt region, surrounding states and from Puerto Rico. The main focus of this study was the corn earworm (*H. zea*) and until 2018 I accumulated data primarily on this host but also detected some evidence of a potentially new nudivirus infecting tobacco budworm (*H. virescens*). In subsequent years, more samples from tobacco budworm and fall armyworm (*Spodoptera* sp.) populations were included.

2.2.1 Sample Collection

A cost-effective and user-friendly method was developed to collect DNA from wild insect populations. In this method, I supplied researchers with sampling kits (Fig. 2.1) that contained prefolded filter papers placed in small individual glassine envelopes, a sampling guide and a sample information form. The sampling kit was designed to be as user-friendly as possible and minimize time and effort needed for tissue collection. Each sampling kit contained six filter papers where insects are placed on (Whatman filter papers, 90mm, Type 43, Ashless – GE Health and Life Sciences), then folded and squeezed to extract the abdominal content into the filter paper. Researcher then removed the cuticular parts leaving only "tissue blots" on the paper. This paper was then folded back and placed into small a glassine envelope (52mm x 90mm, VWR Inc.), which secures the sample from environmental factors and prevented cross contamination. Finally, these sampling kits are shipped to our laboratory via regular mail and upon arrival stored in -20°C freezer until analyzed by PCR.

In addition to filter-paper samples, we also obtained several batches of frozen insects from Louisiana, Mississippi and Kentucky locations. These whole insect samples were processed using a more conventional DNA extraction and precipitation method (Table 2.1) modified from Sambrook & Russell (2006). In this protocol, insect abdomens are dissected and transferred to 1.5 mL microfuge tubes and then incubated for 2 hours at 56°C in alkaline digestion buffer. Then, cellular debris is removed in a salting-out technique and lastly, the DNA is precipitated with ethanol. In addition to frozen insects, we received several high-quality corn earworm DNA extractions prepared from Puerto Rico *H. zea* samples using a phenol-chloroform DNA extraction method. These insects

were collected from 10 different locations in Puerto Rico; Aguadilla (n = 1), Añasco (n = 4), Guánica (n = 24), Guayama (n = 5), Isabela (n = 14), Juana Diaz (n = 5), Lajas (n = 4), Mayagüez (n = 2), Sabana Grande (n = 2), and Santa Isabel (n = 39).

2.2.2 Nudivirus Screening

Hemolymph blotted filter papers were processed by cutting small sections (5mm x 5mm) using sterilized insect dissection scissors and tweezers. Sterilization of dissection utensils was performed by cleaning them with a Kim-wipe and heat-inactivating the remaining nucleic acids in a Bunsen burner for 3 seconds between each sample. The paper section that contained insect tissues (visible due to staining of the paper) was then placed in a 0.6 mL microfuge tube and completely soaked in 0.2 mL TE-buffer (10 mM Tris-CL and 0.1 mM EDTA in distilled water, pH 8.0). TE-buffer was used for DNA extraction (Bereczky et al., 2005) and as a long-term storage medium. In the extraction step, hemolymph-absorbed paper sections were incubated at 98°C for 30 minutes in TE buffer to inhibit enzymatic activities and solubilize the DNA. Finally, suspended particles and contaminants were pelleted via centrifugation at 1000g for 2 minutes and the supernatant was transferred to a new tube to store at -20°C freezer until downstream analyses. Viability of the TE buffer extraction method was tested against a more conventional spin-column method (DNeasy Blood & Tissue Kit, QIAGEN Inc.) with serial dilutions of template DNA source.

A standard polymerase chain reaction (PCR) protocol was used to detect nudivirus sequences in total DNA solutions (Lundberg et al., 1991). PCR reactions were carried out in 10 µL total volume while maintaining the reagent concentrations at the levels specified

by the enzyme manufacturer (Taq DNA Polymerase, recombinant, Thermo Fisher Scientific). We have also tested different annealing temperatures and $MgCl_2$ concentrations using gradient PCR techniques to avoid false negatives and non-specific amplification. Based on these results, the optimal annealing temperature was determined to be 52°C with 2mM $MgCl_2$ final concentration. This $MgCl_2$ concentration was set slightly higher than the suggested value in order to mitigate the chelating effect of EDTA (Ethylenediaminetetraacetic acid) present in TE buffer.

The outcome of a PCR assay can be influenced by several factors, such as primer mismatches, GC-content ratio, and the accuracy of thermal cycler, therefore, the chain reactions were carried out in multiplex setting with two sets of PCR primers (P4 and P13, Lupiani et al., 1999) to minimize false negative errors (Table 2.2). The thermocycler (MJ Research PTC-200) was programmed to run an initial denaturing step at 94°C for 1 minute, 30 rounds of main cycle (annealing temperature: 52°C), and 1 minute of final extension at 72°C. Then, synthesized PCR products (5 µl per sample) were loaded in a 1.5% agarose-TBE (Tris-borate-EDTA) gel which also contains fluorescent nucleic acid stain (GelRed, Biotium). Following electrophoresis (1 hour at 120 V), amplimer banding patterns were visualized using a digital UV trans-illuminator (iBright 1500, Thermo Fisher Scientific). Additionally, a pair of new primers were designed to differentiate *H. zea* from *H. virescens* based on amplimer size (Table 2.2). These primers allow the amplification of ITS1 region, flanked by 18S and 5.8S RNA coding sequences, (Perera et al., 2015) at 58°C annealing temperature. Due to several deletions in *H. virescens* ITS1, the chain reaction generates 448 bp and 413 bp amplimers for *H. zea* and *H. virescens*, respectively.

2.2.3 Data Analyses

In order to assess broad spatial factors, we grouped the sampling locations into three regions; north (IA, IL, IN, NE, OH, SD, UT, WI), mid-south (AR, AZ, KY, NC, OK, TN, TX), and deep south (AL, FL, GA, LA, MS, and SC). Survey regions were compared based on prevalence means using a logistic regression method. Correlation analysis performed on monthly prevalence averages and mean temperature using R statistical software (R Core Team, 2018).

2.2.4 Amplimer Sequencing

Sanger sequencing requires high-quality, high-concentration amplimer solutions. For this purpose, all known nudivirus positives were used as templates for a second round of PCR amplification because multiplex PCR products are not compatible with Sanger sequencing. To purify these DNA samples, proteins, contaminants, and short DNA fragments were removed from the PCR products using magnetic beads (Axygen Prep Mag PCR Cleanup Kit, Corning Inc.) and magnetic separation rack (MagnaRack™, Invitrogen Corp). At this stage, clean PCR products were transferred to the sequencing facility (University of Kentucky Genomics Core Laboratory) in 4 96-well plates where the downstream reactions and the final sequencing steps were carried out. For the main sequencing reaction, P13_Fw primer was used (Table 2.2). Datasets were released in ABI format which contains both sequence and chromatogram information.

Base call quality can be problematic in Sanger sequencing especially at the ends of the amplimers. In our dataset, these low quality bases were removed via QC and trimming workflow in Unipro UGENE program (Okonechnikov et al., 2012). High quality sequences

were then aligned using MUSCLE (Edgar, 2004) program with default settings. The MUSCLE output further trimmed to minimize end gaps and form an alignment block which is necessary for subsequent analyses (Fig. 2.3).

Before proceeding with phylogenetic analysis, an outgroup sequence of 217 bp was generated via Mutation Simulator (https://github.com/mkpython3/Mutation-Simulator) based on the corresponding region of the original HzNV-2 genome. The mutation rate was determined to be 5% of the sequence length. A Bayesian phylogenetic analysis was performed on final alignment block and simulated outgroup sequence using MrBayes (Ronquist et al., 2012) program with total chain length: 1.1M with 10% burn-in, substitution model: GTR, and rate variation: invgamma. The posterior output tree was visualized with FigTree program (http://tree.bio.ed.ac.uk/software/figtree).

2.3 Results

First, we compared our TE buffer-based DNA extraction method with a conventional spin-column kit using serial dilutions of source DNA. This comparison indicated that our extraction method is viable even at low virus titers or DNA concentrations (Fig. 2.2). Furthermore, we noticed that multiplexed PCR has advantages over regular PCR in terms of reducing false negatives.

2.3.1 Summary of the Survey Data

This study was designed to explore HzNV-2 prevalence in the Cotton Belt region and surrounding areas. Field samples were mainly collected between 2016 and 2020, by using pheromone traps, light traps and direct manual collection methods. We initially

focused on *Helicoverpa zea* populations in 2016 and 2017, then included some *H. virescens* and *Spodoptera* samples in the following seasons. The overall sample size was 2291 and this number was achieved with the help of 32 researchers from 21 states ranging from Utah to North Carolina, Texas to Wisconsin (Table 2.3 and 2.6). Molecular screening of the samples showed that 20.81% of all corn earworm (*Helicoverpa zea*) samples were nudivirus infected. By contrast, only 4.3% of the tobacco budworm samples were found to be nudivirus infected. Additionally, samples from four states (AZ, IA, SD and UT) and from other lepidopteran groups (*Spodoptera sp.* and *Cydia pomonella*) showed no evidence on nudivirus presence among analyzed samples. Also, I processed 100 DNA isolates shipped from Puerto Rico in 1.5 mL microfuge tubes which were prepared from corn earworm legs, abdomens and throaces. However, these samples were not included in the *H. zea* nudivirus prevalence dataset due to ambiguities in specimen origin. Our monthly aggregated dataset indicated a significant positive correlation between HzNV-2 prevalence and average monthly temperatures (r_{zea}=0.791, $r_{virescens}$ = 0.692) (Fig. 2.7).

2.3.2 Nudivirus Prevalence in Corn Earworm (*H. zea*) Populations

Among 1403 field collected corn earworm samples, we found 292 nudivirus positives (Fig. 2.5). On average, the nudivirus prevalence was 20.81% in this and it ranged from 4.23% (IN) to 44.44% (OK) (Table 2.3). Since the sample collection pattern was not uniform, inferences about HzNV-2 dynamics were limited. The prevalence means were 13.18%, 22.27%, and 24.22% for the north, mid-south and deep-south regions, respectively (Table 2.4). Two separate comparisons were performed using a logistic regression method. Firstly, I found no significant difference between mid-south and deep south based on

prevalence rates (Df = 11; z = -0.125, p = 0.902). In the second comparison, there was a significant difference in prevalence rates between north and south (mid-south + deep-south) regions (Df = 19; z = -2.273; p = 0.0230). Additionally, aggregated monthly detection graph showed that HzNV prevalence increases starting from April and reaches its peak level in July (Fig. 2.7).

2.3.3 Nudivirus Prevalence in Tobacco Budworm (*H. virescens*) Populations

On average, 4.3% of all *H. virescens* samples (29 positives out of 675 total) showed nudivirus presence (Fig. 2.6). These positives were verified by specific PCR primers (Table 2.2) that distinguish corn earworm from tobacco budworm nudiviruses. The main body of the *H. virescens* samples (540 total) were collected from the Bossier City location (LA) and the nudivirus prevalence was 4.8% (26 positives) in this batch of samples (Table 2.7). A smaller set of samples (111 total) was obtained from Mississippi locations and the nudivirus prevalence was 2.7% (3 positives). Lastly, HvNV was not detected in samples collected from Arkansas locations (24 total).

2.3.4 Puerto Rico Samples

In 2017, I received 100 DNA isolates prepared from *H. zea* thoraces, legs and abdomens. Our initial PCR assay showed that all 24 abdominal DNA samples were nudivirus infected. This result was verified by a subsequent randomized test where 16 randomly selected abdominal DNA isolates are pooled with 16 negative controls. In all tests, abdominal samples yielded strong bands in gel electrophoresis. Also, based on the specimen details provided by the collaborator (Table 2.6), we were not able to determine

the origin of individual DNA sources so these samples excluded from the *H. zea* prevalence analysis due to uncertainties in the dataset. These results indicate that HzNV-2 infection is localized in gonadal tissues rather than causing any detectible systemic infection. These findings are also congruent with the results from previous ultrastructural studies done by Raina & Adams, (1995).

2.3.5 Non-Heliothine Samples

A small portion of the survey dataset was composed of armyworm (*Spodoptera sp.)* and Codling moth (*Cydia pomonella)* samples. Codling moth samples were collected from Oregon (6 total) and none of them showed presence of nudivirus, however, this sample size is too small to make inferences about this species. Moreover, we screened several *Spodoptera* samples; 100 from Tennessee and 6 from North Carolina locations. Similarly, none of the 6 samples from NC showed nudivirus presence but again the sample size was too small for making any inferences. Two out of 100 *Spodoptera* samples from Tennessee were found to be nudivirus positive in the initial PCR assay but yielded very weak bands in gel electrophoresis. These weak bands were not considered to be definitive positives and so were excluded from the dataset.

2.3.6 Amplimer Data

The HzNV-2 positives in this project were sequenced to verify the origin of the sequences and investigate any differences among the sampling locations. Out of 347 nudivirus positives, two *Spodoptera* samples failed to produce adequate amounts of amplimer after the cleaning and amplification steps. Also, sequences from 35 samples were

very short and uninformative. The fragment size of the remaining successful reactions (310 total) ranged from 259 bp to 615 bp. After trimming and quality control process, the read range was 153 bp to 450 bp. These high quality reads were aligned to analyze nucleotide polymorphisms and the average sequence length for this alignment block was 216.28 bp (stdev: 3.95) and the range was 153 bp to 217 bp (Fig. 2.3). Also, pairwise identity of the amplimers in this alignment block was 99.3% with 131 identical sites (60.8%). Analysis of the alignment block resulted in a phylogenetic tree that partially concurs the geographic distribution pattern of all nudivirus positives (Fig. 2.4). Due to high pairwise identity and low numbers of informative variations, phylogenetic relations between multiple nodes remained unresolved.

2.4 Discussion

This project was designed to extend previous survey studies (Lupiani et al., 1999) in both temporal and spatial dimensions. Our results suggest a considerable HzNV-2 prevalence in the Cotton Belt region (24.22%) which is in agreement with Lupiani et al. however I found even higher nudivirus prevalence in Texas (38.57%) where Lupiani et al. did not detect HzNV-2 (Table 2.3). By contrast, I was not able detect viral DNA (n = 6) in our Iowa samples probably due to very small sample size, where Lupiani et al. reported 16.7% incidence rate on average. Also, we found numerous nudivirus positives in locations above the hypothetical overwintering line, such as Nebraska, Wisconsin and Indiana which indicates virus is present in migrating hosts. Finally, HzNV-2 was not detected in samples from four states (UT, AZ, IA and SD) primarily due to small samnpling sizes. This HzNV-

2 prevalence pattern may be attributable to several climatic and host-plant availability parameters however the role of these factors and possible other factors are unknown.

Since multiple research groups were involved in sampling efforts and this was a voluntary effort, it was impossible to achieve a consistent sampling regime. To mitigate this issue, we partitioned the dataset into three regions; deep-south, mid-south and north, based on climatic features and host plant availability. One-way ANOVA test revealed that these regions were significantly different in terms of virus prevalence means, and pairwise comparisons showed significant difference between the north region and other regions (Table 2.4). Statistical tests show that HzNV-2 prevalence in migrating populations (northern) are significantly lower than that of overwintering populations (Deep-south and Mid-south). In southern overwintering populations, HzNV-2 appears around 20^{th} and 23^{rd} weeks of the year (deep south and mid-south, respectively) but in northern populations, the virus first appears around the 26^{th} week with the appearance of migrating populations from the south.

The correlation between prevalence and temperature occurs potentially due to increasing host population density. (Fig. 2.7). In addition to climatic factors, prevailing winds play an important role in Heliothine dispersal. In North America, March and April are the windiest months and in this time frame, northerly prevailing winds become more southerly winds (Ward, 1916; Westbrook & López, 2010; Jones et al., 2019). This transition in prevailing wind regime in the spring months can be the primary factor driving recolonization of non-overwintering regions. In conjunction with these prevailing winds, climatic pressure gradients are also effective in the cool layer of air that forms at night adjacent to the ground (nocturnal boundary layer) (Drake, 1984; Drake & Farrow, 1988;

Lingren et al., 1995; Sandstrom et al., 2007) where macro-insects use most for dispersal and migration.

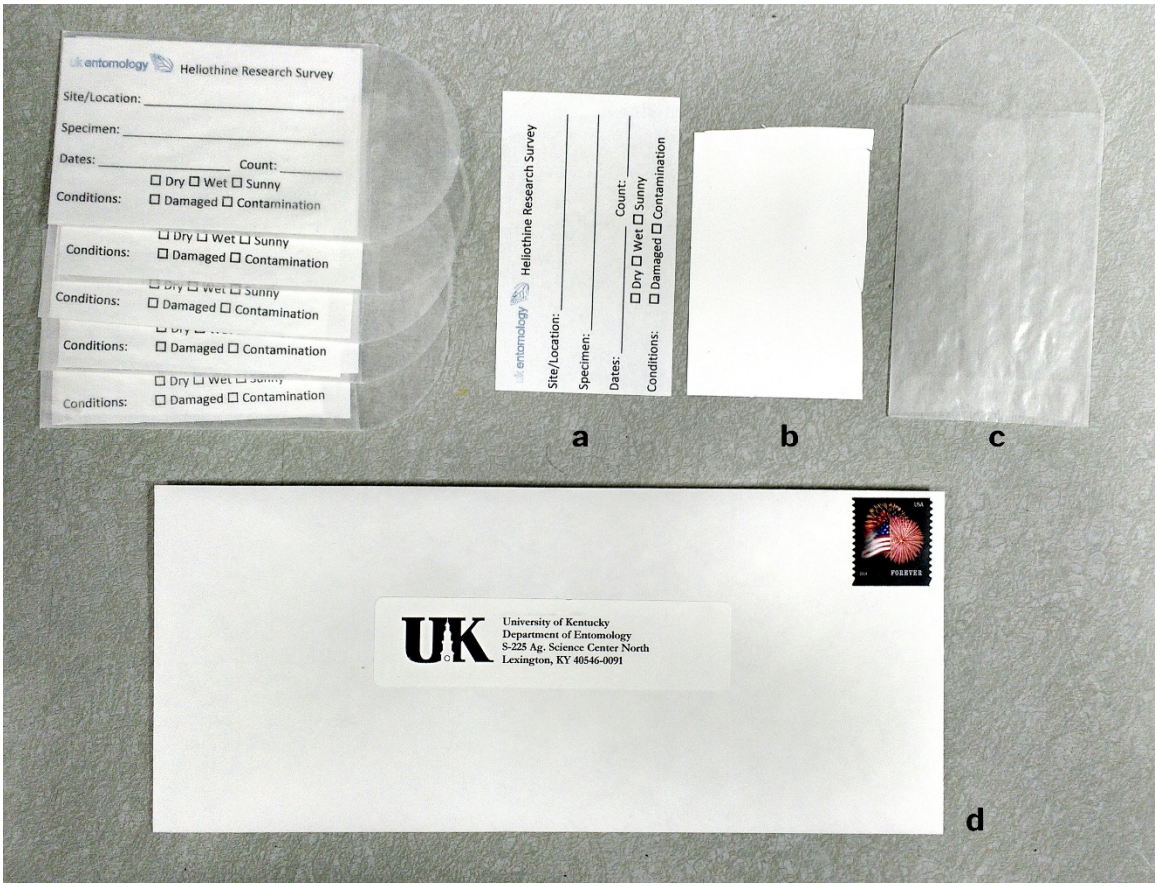
Analysis of the Puerto Rico dataset indicated that the HzNV-2 can only be detected from abdominal tissues since no virus DNA was detected from thorax and leg tissues. Even though this result is predictable from virus pathology, further studies with greater statistical power are needed to verify this phenomenon. As these samples were collected and prepared by professionals, from different host plants in different time frames, cross contamination is unlikely to be an issue for this dataset

Another significant outcome of this survey is the evidence of a novel nudivirus species detected in *H. virescens* populations. The host species identification was verified using a specific primer set (Table 2.2) that targets a deletion in the *H. virescens* host genome (Gilligan et al., 2015). These assays indicate that a novel nudivirus is circulating in southeastern *H. virescens* populations with mean 4.2% incidence rate. Moreover, DNA sequencing and phylogenetic analysis of the amplimer dataset showed a significant divergence between HzNV-2 and *H. virescens* nudivirus (Fig. 2.4). Further information on genomic features and secondary haplotypes are presented in Chapter 4. However, all the findings represented here do not provide sufficient data to fully describe the regional origins, transmission and interaction of this new virus with its *H. virescens* host. Once isolated and propagated in the laboratory, the *H. virescens* nudivirus can be investigate for its pathology and potential as a biopesticide.

Finally, my findings serve as a starting point for future HzNV-2 researchers, field professionals and developers of bioinsecticides. The sampling method we developed for this study can be used for other insect taxa to study novel entomopathogens. Since the wet

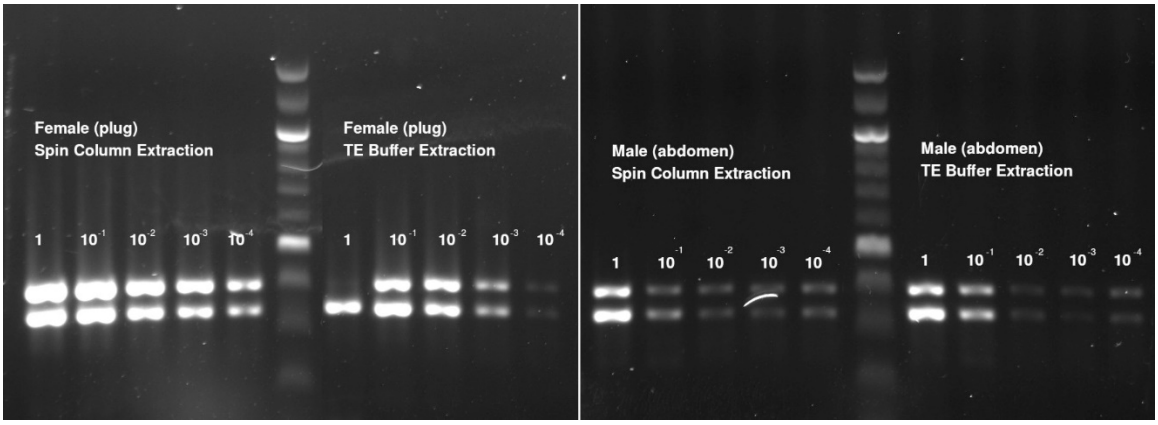
laboratory techniques and the workflow described here are safe and cost effective, it can be used for educational purposes or can be used safely to train medical/forensic personnel and may help citizen scientists to contribute similar research projects to extend sample collection efforts.

Figure 2.1: Overview of the sampling kit



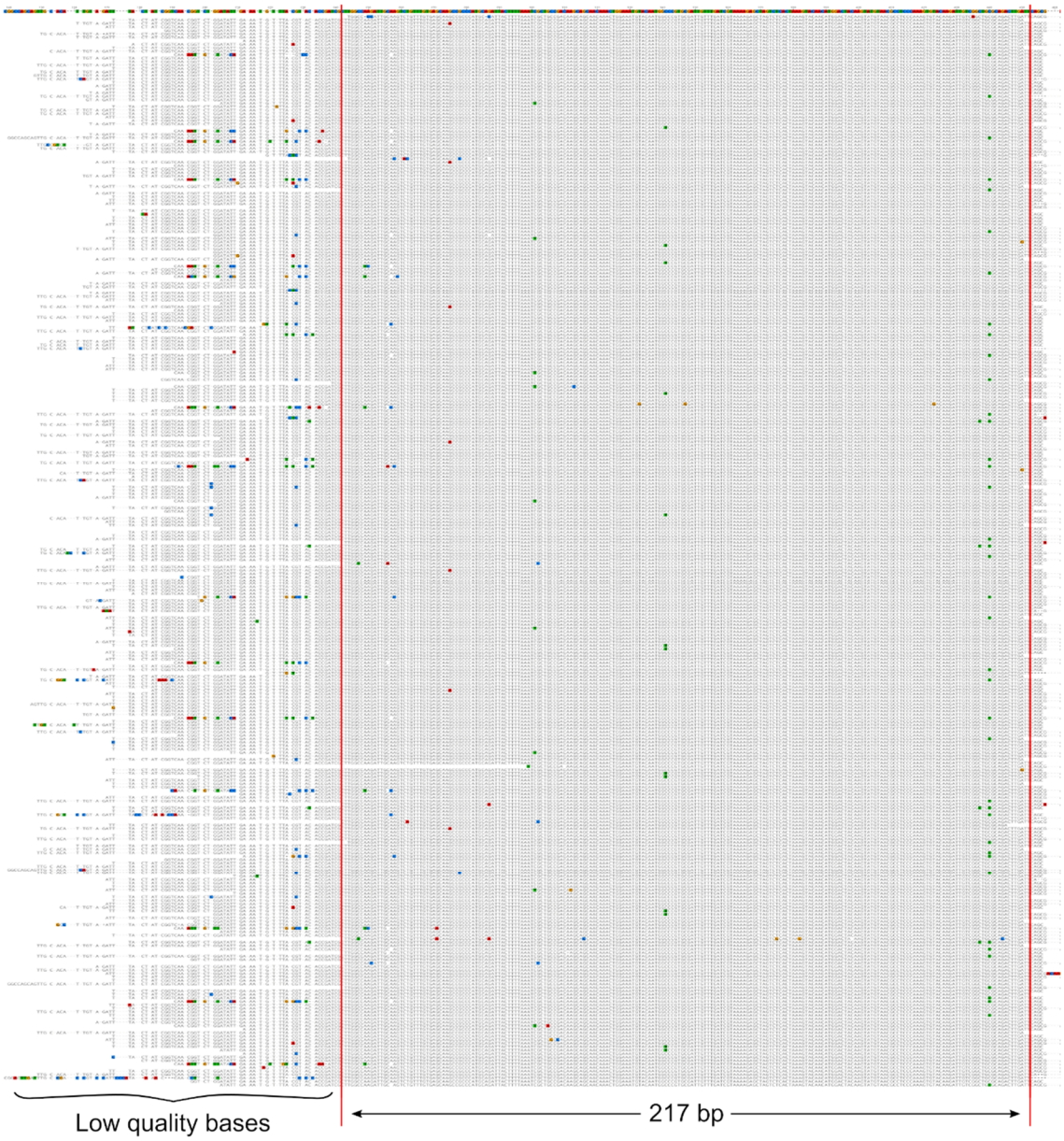
Field samples were collected using the sampling kit that contains information sheet (a), filter paper for abdominal tissue blotting (b), and a glassine envelope for preventing cross-contamination (c). Sampling kits were returned via regular mail service (d).

Figure 2.2: PCR sensitivity assay: Comparison of PCR results that are produced using standard spin column extraction method (left) and TE buffer extraction method (right).



The multiplexed PCR method for screening field-collected samples uses two sets of primers. On the left panel, PCR with undiluted TE buffer extraction of female plug failed with P13 primers however multiplexed reaction helped to prevent false negative errors.

Figure 2.3: Multiple sequence alignment of 310 nudivirus amplimer sequences.



SNPs and deletions are indicated in colors and spaces. Low quality bases were trimmed using UGENE trimming workflow. First line in the list shows the consensus sequence.

Figure 2.4: Summary of the phylogenetic relations among the amplimer sequences. A 350 bp section of all nudivirus samples were sequenced and 217 bp alignment block was used to build the phylogenetic tree.

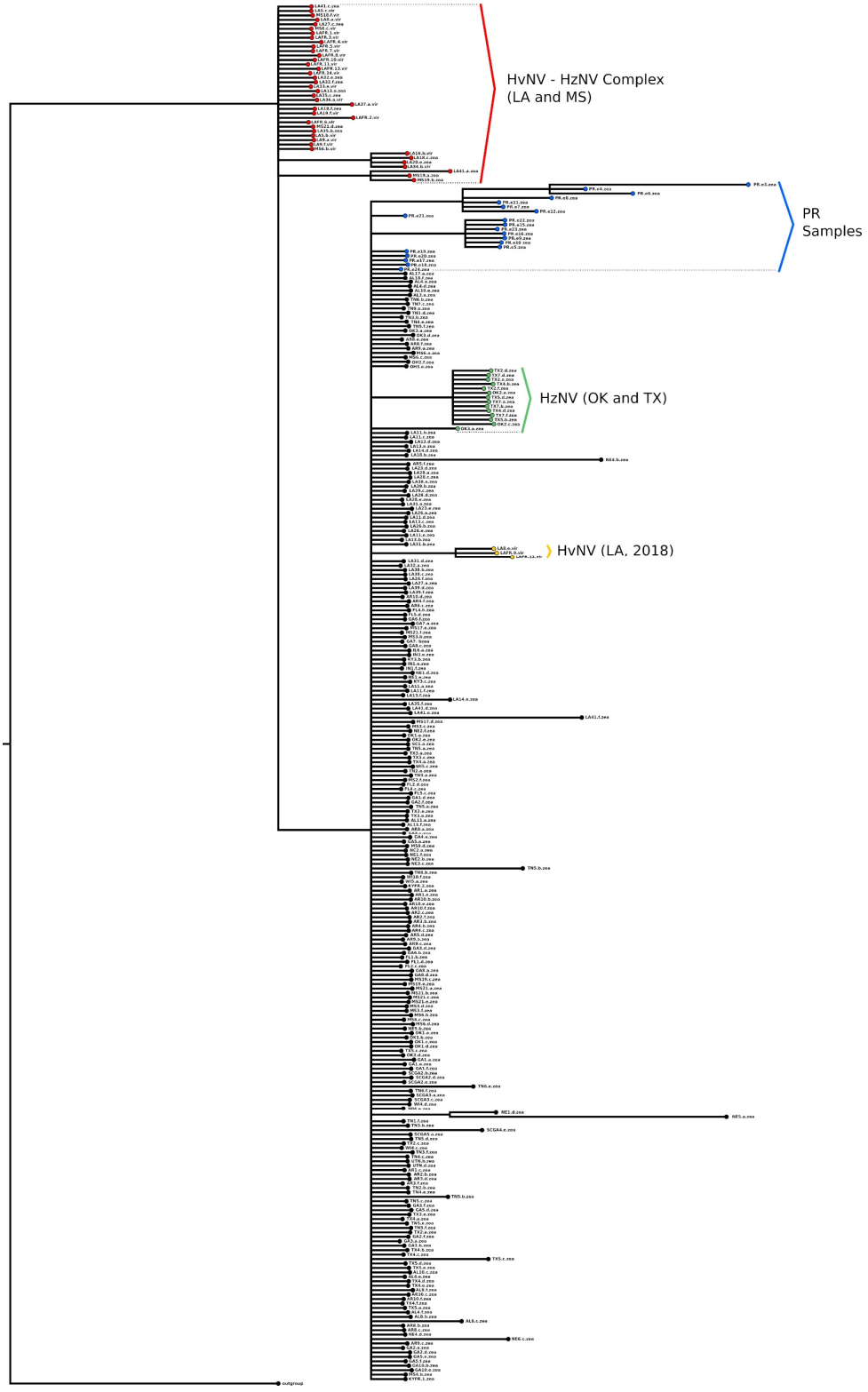
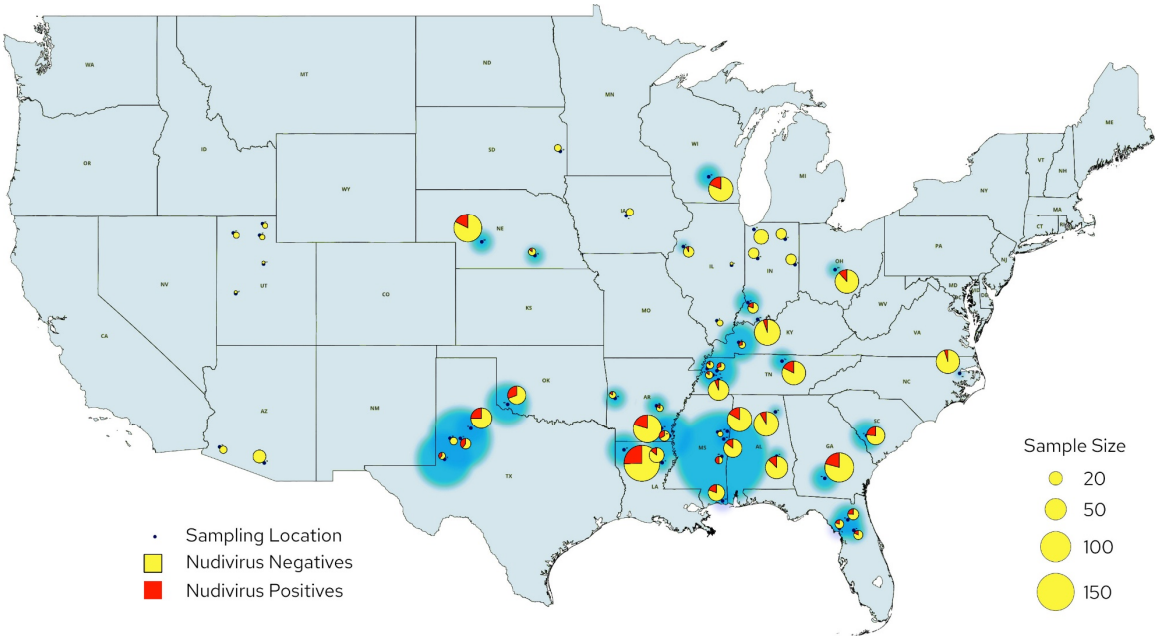
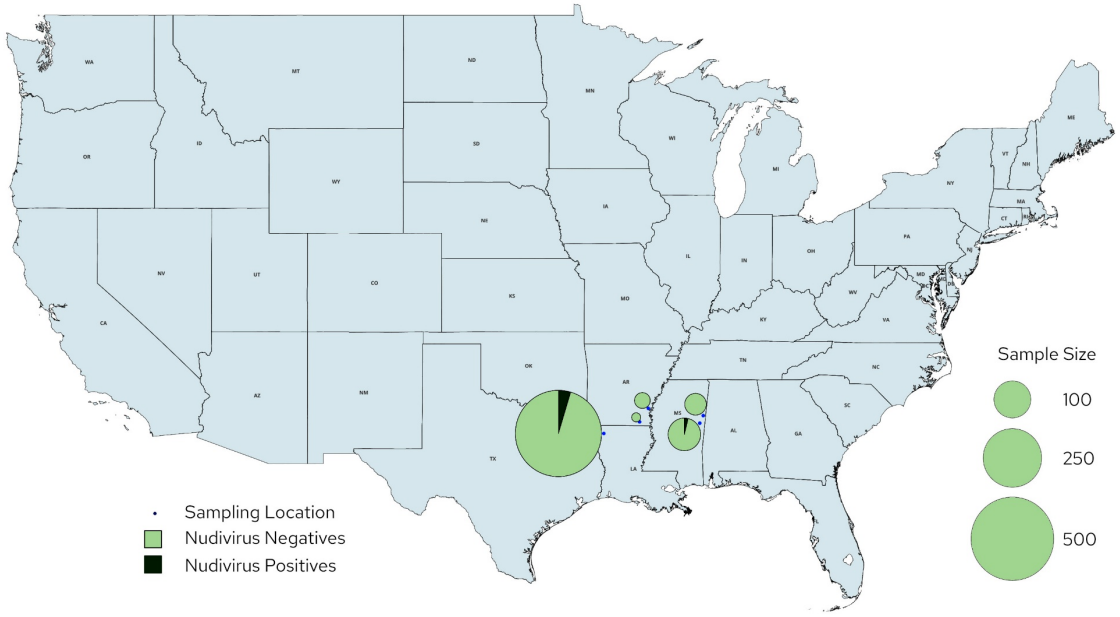


Figure 2.5: HzNV detections and underlying prevalence heatmap in feral *H. zea* populations



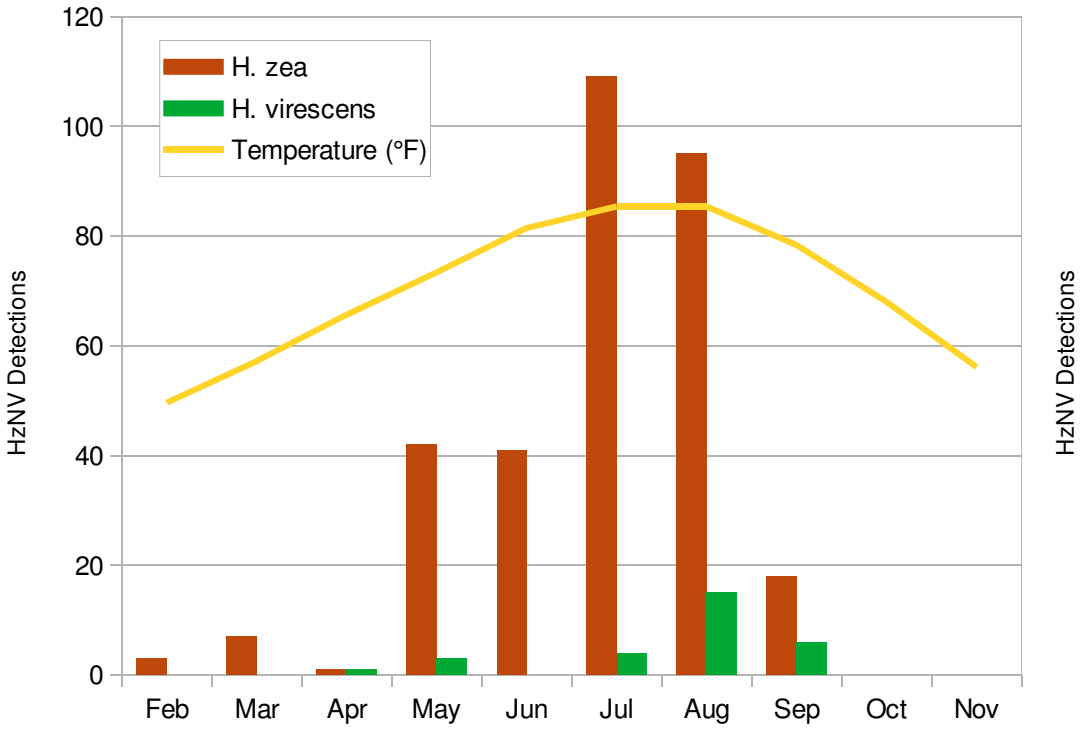
A total of 1403 samples were collected from H. zea populations. PCR testing of these samples revealed 292 HzNV-2 positives with 20.81% prevalence on average. The area of the blue discs is correlated with the associated prevalence rates.

Figure 2.6: HvNV detections and underlying prevalence heatmap in feral *H. virescens* populations.



A total of 675 samples were collected from H. zea populations. PCR test of these samples showed 29 nudivirus positives with 4.3% prevalence on average.

Figure 2.7: Aggregated monthly HzNV detection in corn earworm (*H. zea*) and tobacco budworm (*H. virescens*) populations with monthly temperature averages.



A positive correlation between HzNV prevalence and average monthly temperatures was detected (r_{zea}=0.791, $r_{virescens}$ = 0.692). This correlation is also influenced by host biology and population density.

Table 2.1: Protocol for DNA extraction from whole insect samples.

Reagents and Materials

Digestion Buffer (pH 8.0)*	Chilled Ethanol (70% and 98%)
Proteinase K (20 mg/mL)	Pipettes and tips
Sodium Acetate 3M (pH 5.2)	1.5 mL microcentrifuge tubes

*per 100 mL = 92.14 µL ddH20 + 2mL EDTA + 0.66 mL Tris + 5mL SDS + 0.2 mL NaCl

Equipment

- Heat block, Centrifuge, and Vortex

Procedure

Tissue Digestion

- Add 5 µL Proteinase K to each 200 µL of Digestion Buffer (final 0.5mg/mL)
- Place pieces of the abdomen in each of the solutions. Sanitize tools in between each.
- Place in the heat block for 25 minutes at 55 ^{o}C.
- Vortex and place in the centrifuge at max speed for 20 minutes
- Transfer Supernatant into the new tube

Precipitation of Protein and Cell Debris

- Add 1/10 volume of Sodium Acetate 3M pH 5.2 (final 0.3M)
- Invert to mix and put in a -20 degree celsius for 15 minutes
- Centrifuge at max speed for 20 minutes
- Transfer 300 µL of supernatant to a new tube

Precipitation of Nucleic Acids

- Add 600 µL of 98% ethanol (final 60-80%)
- Mix gently by inverting and incubate for 20 minutes at -20 ^{o}C
- Centrifuge at max speed for 20 minutes
- Wash pellet with ethanol and discard supernatant
- Place pellet in heat block until dry
- Resuspend in nuclease-free water

This protocol was modified from Sambrook & Russell, 2006

Table 2.2: List of PCR primers used for screening HzNV-2 presence in field-collected samples.

Name	Direction	Sequence	Amplimer Size
P4	Forward	GCACGATTCCGTAATGTTC	404 bp
P4	Reverse	GCACACCTATCAATCACC	
P13	Forward	TCGATGCCGTAATACC	320 bp
P13	Reverse	GTCGCTGAATCAAGTCTG	
Hvir	Forward	GAACCTGCGGAAGGATCATTAAC	*H. zea*: 448 bp, *H. virescens*: 413 bp
Hvir	Reverse	CCGTCCAGGGTAATAGTTTTAAATTAC	

Annealing temperature for P4, P13 (Lupiani et al., 1999) and Hvir are 52°C, 52°C and 58°C respectively. Hvir primers amplify an internal transcriber spacer (ITS1) flanked by 18S and 5.8S RNA encoding sequences. A 35 bp deletion in *H. virescens* ITS1 is distinguishable from *H. zea* ITS1.

Table 2.3: Summary of the HzNV incidence in *H. zea* populations.

State	County	Nudivirus Positive	Total Samples	Prevalence (%)
AL	DeKalb	6	66	9.09
	Shorter	8	54	14.81
AR	Chicot	6	12	50.00
	Lonoke	22	86	25.58
	Camden	1	5	20.00
	Scott	1	5	20.00
AZ	Pima	0	18	0.00
	Yuma	0	6	0.00
FL	Alachua	4	13	30.77
	Levy	2	8	25.00
	Marion	2	9	22.22
GA	Tift	24	91	26.37
IA	Story	0	6	0.00
IL	Champaign	0	1	0.00
	Franklin	0	4	0.00
	Warren	1	12	8.33
IN	Knox	3	12	25.00
	Porter	0	23	0.00
	Randolph	0	12	0.00
	Tippecanoe	0	12	0.00
	Whitley	0	12	0.00
KY	Caldwell	2	5	40.00
	Fayette	4	72	5.56
LA	Bossier	47	138	34.06
	Franklin	4	24	16.67
MS	Lowndes	13	65	20.00
	Jackson	7	29	24.14
	Lauderdale	6	6	100.00
	Noxubee	6	36	16.67
	Oktibbeha	0	3	0.00
NC	Washington	3	60	5.00
NE	Clay	1	6	16.67
	Lincoln	18	84	21.43
OH	Columbus	8	60	13.33
OK	Jackson	16	36	44.44

Table 2.3 (continued)

SC	Barnwell	11	36	30.56
SD	Brookings	0	5	0.00
TN	Crockett	1	6	16.67
	Dyer	1	6	16.67
	Gibson	3	7	42.86
	Jackson	13	60	21.67
	Madison	3	46	6.52
TX	Andrews	3	5	60.00
	Crosby	16	47	34.04
	Lynn	8	12	66.67
	Terry	0	6	0.00
UT	Box Elder	0	4	0.00
	Cache	0	3	0.00
	Millard	0	1	0.00
	Utah	0	1	0.00
	Weber	0	3	0.00
WI	Columbia	15	64	23.44
Total		292	1403	20.81

Table 2.4: HzNV prevalence in North, Mid-South and Deep South regions with logistic regression results

Region	Nudivirus Positives	Sample Size	Prevalence (%)
North	46	349	13.18
Survey Average	*292*	*1403*	*20.81*
Mid-South	106	476	22.27
Deep South	140	578	24.22

North states: IA, IL, IN, NE, OH, SD, UT, and WI.
Mid-south states: AR, AZ, KY, NC, OK, TN, and TX.
Deep south states: AL, FL, GA, LA, MS, and SC.

Two separate comparisons were performed using logistic regression method. In the first comparison, no significant difference between mid-south and deep-south was detected based on prevalence rates (Df = 11; z = -0.125, p = 0.902). In the second comparison, there was a significant difference in prevalence rates between north and south (mid-south + deep-south) regions (Df = 19; z = -2.273; p = 0.0230)

Table 2.5: Concatenated weekly table of HzNV-2 prevalence and incidence in *H. zea* populations

		W19	W20	W21	W22	W23	W24	W25	W26	W27	W28	W29	W30	W31	W32	W33	W34	W35	W36	W37	W38	W40
North	IA														0 (6)							
	IL								0 (5)				0.2 (1/5)			0 (7)						
	IN								0.17 (3/18)		0 (1)	0 (5)	0 (8)	0 (12)	0 (3)	0 (12)	0 (12)					
	OH										0.25 (¼)	0.38 (3/8)	0 (1)	0.17 (1/6)	0.1 (1/10)	0.05 (1/20)	0.09 (1/11)					
	SD																0 (5)					
	UT										0 (4)				0 (8)							
	WI								0 (1)	0.17 (1/6)	0.33 (2/6)		0.33 (1/3)	0.4 (4/10)	0.18 (2/11)	0 (10)	0.36 (4/11)	0.2 (1/5)	0 (1)			
Mid-South	AR					0.17 (1/6)	0.33 (2/6)	0.06 (1/18)	0.17 (1/6)		0.29 (7/24)	0.44 (8/18)	0.42 (5/12)	0.42 (5/12)	0 (6)							
	KY							0 (1)			0 (1)					0.67 (2/3)				0.06 (4/72)		
	NC									0 (3)		0.33 (1/3)	0 (6)		0 (6)	0.17 (1/6)	0 (6)	0 (12)	0 (6)			0.08 (1/12)
	OK												0.5 (3/6)		0.5 (6/12)	0.5 (4/8)	0.25 (¼)	0.33 (2/6)				
	TN		0 (6)		0 (3)						0 (5)			0.3 (3/10)	0.2 (2/10)	0.2 (6/30)	0.25 (6/24)	0.06 (1/17)	0.1 (1/10)	0.2 (2/10)		
	TX						0.33 (4/12)	0 (6)	0.33 (2/6)	0.6 (3/5)	0.64 (7/11)	0.17 (2/12)		0.55 (6/11)	0.86 (6/7)							
Deep South	AL		0.17 (3/18)	0.06 (1/18)	0.21 (5/24)						0 (6)	0.06 (1/18)	0.11 (2/18)	0 (6)		0.17 (2/12)						
	FL				0.5 (2/4)			0 (3)	0.5 (2/4)				0 (3)	0 (2)		0.29 (4/14)						
	GA									0 (7)	0.5 (6/12)	0.42 (5/12)	0.33 (4/12)	0.33 (2/6)	0.33 (4/12)	0.08 (1/12)	0.08 (1/12)	0.17 (1/6)				
	LA		0.17 (2/12)		0 (6)	0.33 (4/12)	0.17 (2/12)	0.67 (4/6)	0.5 (6/12)	0.5 (3/6)	0.31 (4/13)	0.4 (2/5)	0 (6)	0.25 (3/12)	0.67 (4/6)	0.83 (5/6)						
	MS		0 (6)	0.33 (2/6)	0 (6)	0 (6)			0 (9)	0 (12)	0.25 (3/12)	0.64 (7/11)	0 (9)	0.5 (6/12)	0.67 (6/9)	0.08 (1/12)	0.11 (1/9)	0 (8)		0.5 (6/12)		
	SC							0.17 (1/6)		0.25 (3/12)	0.17 (1/6)	0.5 (3/6)	0.5 (3/6)									

Color intensity of each cell is correlated with prevalence rate and numbers in parentheses show HzNV positives (if applicable) over sample size. Cells without numbers denotes no data available.

Table 2.6: Summary of the Puerto Rico dataset

Tube ID	Species	Identification method	Locality	Date collected	Latitude	Longitude	Crop	DNA source
t4	*H. zea*	genital morphology	Guánica	24-Aug-16	17.98633	-66.90435	pepper	thorax
t5	*H. zea*	genital morphology	Guayama	24-Aug-16	17.98781	-66.21546	soybean	thorax
t7	*H. zea*	genital morphology	Guayama	24-Aug-16	17.98781	-66.21546	soybean	thorax
t8	*H. zea*	genital morphology	Guayama	24-Aug-16	17.98781	-66.21546	soybean	thorax
t10	*H. zea*	genital morphology	Añasco	24-Feb-16	18.27339	-67.15700	corn	leg
t11	*H. zea*	genital morphology	Añasco	24-Feb-16	18.27470	-67.15669	pepper	leg
t12	*H. zea*	genital morphology	Añasco	9-Mar-16	18.27339	-67.15700	corn	leg
t13	*H. zea*	genital morphology	Añasco	9-Mar-16	18.27470	-67.15669	pepper	leg
t14	*H. zea*	genital morphology	Aguadilla	24-Feb-16	18.44705	-67.12185	pigeonpea pepper	leg
t15	*H. zea*	genital morphology	Sabana Grande	22-Feb-16	18.08432	-66.94881	pigeonpea	leg
t16	*H. zea*	genital morphology	Sabana Grande	7-Mar-16	18.08432	-66.94881	pigeonpea	leg
t17	*H. zea*	genital morphology	Lajas	4-Apr-16	18.03219	-67.07036	corn	leg
t18	*H. zea*	genital morphology	Lajas	2-May-16	18.03189	-67.07178	corn/Phaseolus	leg
t19	*H. zea*	genital morphology	Lajas	16-May-16	18.03189	-67.07178	corn/Phaseolus	leg
t20	*H. zea*	genital morphology	Lajas	14-Jun-16	18.03189	-67.07178		leg
t21	*H. zea*	genital morphology	Mayagüez	19-May-16	18.21997	-67.1443	corn	leg
t22	*H. zea*	genital morphology	Mayagüez	7-Oct-16	18.21991	-67.14683	tomato/corn	leg
t23	*H. zea*	genital morphology	Isabela	24-Feb-16	18.46305	-67.05431	corn	leg
t24	*H. zea*	genital morphology	Isabela	9-Mar-16	18.47076	-67.04997	soybean	leg
t25	*H. zea*	genital morphology	Isabela	4-Apr-16	18.46653	-67.04546	crotalaria	leg
t26	*H. zea*	genital morphology	Isabela	5-May-16	18.47076	-67.04997	soybean	leg
t27	*H. zea*	genital morphology	Isabela	31-May-16	18.47076	-67.04997		leg
t28	*H. zea*	genital morphology	Isabela	13-Jul-16	18.47076	-67.04997	sorghum	leg
t29	*H. zea*	genital morphology	Isabela	31-Aug-16	18.47076	-67.04997		leg
t30	*H. zea*	genital morphology	Isabela	31-Aug-16	18.47093	-67.04815	phaseolus	leg
t31	*H. zea*	genital morphology	Isabela	7-Sep-16	18.47076	-67.04997		leg
t32	*H. zea*	genital morphology	Isabela	7-Sep-16	18.47076	-67.04997		leg
t33	*H. zea*	genital morphology	Isabela	22-Sep-16	18.47076	-67.04997		leg
t34	*H. zea*	genital morphology	Isabela	25-Oct-16	18.47076	-67.04997		leg
t35	*H. zea*	genital morphology	Juana Diaz	30-Mar-16	18.02813	-66.52991	phaseolus	leg
t36	*H. zea*	genital morphology	Juana Diaz	30-Mar-16	18.02813	-66.52991	phaseolus	leg
t37	*H. zea*	genital morphology	Juana Diaz	21-Oct-16	18.03151	-66.52896	phaseolus	leg
t38	*H. zea*	genital morphology	Juana Diaz	4-Nov-16	18.03151	-66.52896		leg
t39	*H. zea*	genital morphology	Guánica	22-Feb-16	18.00935	-66.89254	pigeonpea sunflower	leg
t40	*H. zea*	genital morphology	Guánica	7-Mar-16	17.97986	-66.90089	pepper	leg
t41	*H. zea*	genital morphology	Guánica	4-Apr-16	17.99534	-66.96012	soybean	leg
t42	*H. zea*	genital morphology	Guánica	2-May-16	18.07454	-66.96381	pepper	leg
t43	*H. zea*	genital morphology	Guánica	16-May-16	17.97986	-66.90089		leg
t44	*H. zea*	genital morphology	Guánica	31-May-16	17.97986	-66.90089		leg

Table 2.6 (continued)

t45	*H. zea*	genital morphology	Guánica	14-Jun-16	17.97986	-66.90089		leg
t46	*H. zea*	genital morphology	Guánica	28-Jun-16	17.97986	-66.90089		leg
t47	*H. zea*	genital morphology	Guánica	16-Jul-16	17.97986	-66.90089		leg
t48	*H. zea*	genital morphology	Guánica	26-Jul-16	17.97986	-66.90089		leg
t49	*H. zea*	genital morphology	Guánica	10-Aug-16	17.97986	-66.90089		leg
t50	*H. zea*	genital morphology	Guánica	7-Sep-16	17.97986	-66.90089		leg
t51	*H. zea*	genital morphology	Guánica	21-Sep-16	17.97986	-66.90089		leg
t52	*H. zea*	genital morphology	Guánica	7-Oct-16	17.97986	-66.90089		leg
t53	*H. zea*	genital morphology	Guánica	19-Oct-16	17.97986	-66.90089		leg
t54	*H. zea*	genital morphology	Guánica	4-Nov-16	17.97986	-66.90089		leg
t55	*H. zea*	genital morphology	Santa Isabel	22-Feb-16	17.98297	-66.41855	tomato	leg
t56	*H. zea*	genital morphology	Santa Isabel	22-Feb-16	18.03792	-66.39095	pepper	leg
t57	*H. zea*	genital morphology	Santa Isabel	7-Mar-16	17.98297	-66.41855	tomato	leg
t58	*H. zea*	genital morphology	Santa Isabel	7-Mar-16	17.99873	-66.41795	tomato	leg
t59	*H. zea*	genital morphology	Santa Isabel	4-Apr-16	17.98297	-66.41855	tomato	leg
t60	*H. zea*	genital morphology	Santa Isabel	4-Apr-16	17.99873	-66.41795	tomato	leg
t61	*H. zea*	genital morphology	Santa Isabel	2-May-16	17.99769	-66.42598	tomato	leg
t62	*H. zea*	genital morphology	Santa Isabel	2-May-16	17.99769	-66.42598	tomato	leg
t63	*H. zea*	genital morphology	Santa Isabel	9-May-16	17.99873	-66.41795	tomato	leg
t64	*H. zea*	genital morphology	Santa Isabel	9-May-16	17.98287	-66.43085	pepper	leg
t65	*H. zea*	genital morphology	Santa Isabel	16-May-16	17.99756	-66.42691	tomato	leg
t66	*H. zea*	genital morphology	Santa Isabel	16-May-16	17.99756	-66.42691	tomato	leg
t67	*H. zea*	genital morphology	Santa Isabel	31-May-16	17.98287	-66.43085		leg
t68	*H. zea*	genital morphology	Santa Isabel	14-Jun-16	17.98547	-66.42914	pepper	leg
t69	*H. zea*	genital morphology	Santa Isabel	14-Jun-16	17.98547	-66.42914	pepper	leg
t70	*H. zea*	genital morphology	Santa Isabel	28-Jun-16	17.98547	-66.42914	pepper	leg
t71	*H. zea*	genital morphology	Santa Isabel	28-Jun-16	17.98547	-66.42914	pepper	leg
t72	*H. zea*	genital morphology	Santa Isabel	12-Jul-16	17.98547	-66.42914	pepper	leg
t73	*H. zea*	genital morphology	Santa Isabel	12-Jul-16	17.98547	-66.42914	pepper	leg
t74	*H. zea*	genital morphology	Santa Isabel	28-Jul-16	17.98547	-66.42914	pepper	leg
t75	*H. zea*	genital morphology	Santa Isabel	10-Aug-16	17.98547	-66.42914	pepper	leg
t76	*H. zea*	genital morphology	Santa Isabel	24-Aug-16	17.98547	-66.42914	pepper	leg
t77	*H. zea*	genital morphology	Santa Isabel	7-Sep-16	17.98547	-66.42914	pepper	leg
t78	*H. zea*	genital morphology	Santa Isabel	7-Oct-16	17.98547	-66.42914	pepper	leg
t79	*H. zea*	genital morphology	Santa Isabel	4-Nov-16	17.98547	-66.42914		leg
t80	*H. zea*	genital morphology	Guayama	7-Sep-16	17.98719	-66.21555	soybean	leg
t81	*H. zea*	genital morphology	Guayama	21-Sep-16	17.98719	-66.21555	soybean	leg
e1	*H. zea*	Real Time PCR ITS1	Isabela	17-Feb-16	18.47076	-67.04997	soybean	abdomen
e2	*H. zea*	Real Time PCR ITS1	Isabela	20-Apr-16	18.46653	-67.04546	crotalaria	abdomen
e3	*H. zea*	Real Time PCR ITS1	Santa Isabel	24-Feb-16	17.98297	-66.41855	tomato	abdomen

Table 2.6 (continued)

e4	*H. zea*	Real Time PCR ITS1	Santa Isabel	7-Mar-16	17.98297	-66.41855	tomato	abdomen
e5	*H. zea*	Real Time PCR ITS1	Santa Isabel	7-Mar-16	17.98297	-66.41855	tomato	abdomen
e6	*H. zea*	Real Time PCR ITS1	Santa Isabel	7-Mar-16	17.99873	-66.41795	tomato	abdomen
e7	*H. zea*	Real Time PCR ITS1	Guánica	22-Feb-16	18.00935	-66.89254	pigeonpeas sunflower	abdomen
e8	*H. zea*	Real Time PCR ITS1	Santa Isabel	7-Mar-16	17.98297	-66.41855	tomato	abdomen
e9	*H. zea*	Real Time PCR ITS1	Santa Isabel	7-Mar-16	17.99873	-66.41795	tomato	abdomen
e10	*H. zea*	Real Time PCR ITS1	Santa Isabel	7-Mar-16	17.99873	-66.41795	tomato	abdomen
e11	*H. zea*	Real Time PCR ITS1	Guánica	2-May-16	18.00935	-66.89254	pigeonpeas unflower	abdomen
e12	*H. zea*	Real Time PCR ITS1	Guánica	28-Jun-16	17.98633	-66.90435	pepper	abdomen
e13	*H. zea*	Real Time PCR ITS1	Santa Isabel	2-May-16	17.99769	-66.42598	tomato	abdomen
e14	*H. zea*	Real Time PCR ITS1	Juana Diaz	7-Mar-16	18.02813	-66.52991	phaseolus	abdomen
e15	*H. zea*	Real Time PCR ITS1	Santa Isabel	2-May-16	17.99873	-66.41795	tomato	abdomen
e16	*H. zea*	Real Time PCR ITS1	Santa Isabel	24-May-16	17.98297	-66.41855	tomato	abdomen
e17	*H. zea*	Real Time PCR ITS1	Guánica	31-May-16	17.98633	-66.90435	pepper	abdomen
e18	*H. zea*	Real Time PCR ITS1	Guánica	31-May-16	17.98615	-66.90225	tomato	abdomen
e19	*H. zea*	Real Time PCR ITS1	Guánica	31-May-16	17.98615	-66.90225	tomato	abdomen
e20	*H. zea*	Real Time PCR ITS1	Santa Isabel	31-May-16	17.98493	-66.42953	pepper	abdomen
e21	*H. zea*	Real Time PCR ITS1	Santa Isabel	31-May-16	17.98493	-66.42953	pepper	abdomen
e22	*H. zea*	Real Time PCR ITS1	Santa Isabel	31-May-16	17.98547	-66.42914	pepper	abdomen
e23	*H. zea*	Real Time PCR ITS1	Santa Isabel	31-May-16	17.98369	-66.42665	pepper	abdomen
e24	*H. zea*	Real Time PCR ITS1	Guánica	31-May-16	18.00745	-66.88863	pepper	abdomen

Table 2.7: Summary of the nudivirus incidence in *H. virescens* populations.

State	County	NV Positives	Total Samples	Prevalence
AR	Ashley	0	6	0
	Desha	0	18	0
LA	Bossier	26	540	4.81%
MS	Lowndes	0	34	0
	Noxubee	3	77	3.89%
Total		29	675	4.3%

CHAPTER 3: DIGITAL SURVEY OF HELICOVERPA ZEA NUDIVIRUS

3.1 Introduction

An unprecedented amount of genetic data has accumulated in public databases and it is increasing daily. This presents many opportunities for new approaches to investigate biology, and in particular, pathogens associated with large data sets. The NCBI database contains more than 1.8 billion genetic sequences and 24.7 million of them were deposited between June 2021 and August 2021 (ncbi.nlm.nih.gov/genbank/statistics). Modern next-generation sequencing platforms (NGS) generate vast amount of read data in an untargeted fashion. Most of these platforms use microfluidic flow cells to hybridize DNA fragments on to an adapter-coated glass slide and perform sequencing by synthesis. In general, flow cells are designed to generate millions of reads in a massively parallelized manner. Some library preparation methods involve a random DNA fragmentation step in which the entire genetic material is processed and fragmented haphazardly. This randomized and massively parallelized sequencing strategy enables reconstruction of genomic and transcriptomic composition and further investigation of variation discovery and pathological diagnosis (Lipkin, 2013; Parize et al., 2017).

The term big data refers to a paradigm shift in which the data generated with these new technologies overwhelms the methods conventionally used. The average data size generated by a NGS system is larger - by orders of magnitude, compared to previous generation Sanger sequencing systems. Naturally, these larger datasets require significant amounts of computing power and network resources to obtain the raw data from the database and perform the statistical analyses. Recent advances in data processing technologies and the introduction of high-power computing resources to a broader range

of analysts allowed researchers to run many resource demanding algorithms in a time and cost-effective manner. Data mining is the process of probing large datasets to find patterns and valuable information by using specific software and analytic tools. Medical data mining has been a standard procedure since the discovery of DNA sequencing in the early 1970s. In the following decades, introduction of low cost, high-throughput sequencing and data mining applications provide new approaches to investigate infectious diseases (Goldberg et al., 2015; Lecuit & Eloit, 2015) and they have been widely employed during CoVID-19 pandemic (Kumar Das et al., 2021). Here, in this chapter, I use data mining methods to investigate nudivirus presence in public sequence databases.

Insect viruses are a diverse group of entomopathogens with nearly 1200 known strains that infect more than 20 insect families (Grzywacz, 2017). An understudied group of entomopathogens, the Nudiviruses, were formally classified as a non-occluded Baculovirus subfamily as they don't routinely form proteinaceous inclusions or, polyhedral bodies, that are typical of the baculoviruses, entomopoxviruses, granuloviruses and cypoviruses. Comparative genomics showed that two insect virus families, the Nudiviridae and the Baculoviridae, evolved from a common ancestor nearly 310 million years ago (Thézé et al., 2011). These two large dsDNA virus families share at least 15 homologous core genes (Wang, et al., 2007) but can differ significantly in terms of pathology and life cycle. Despite the taxonomic uncertainties in the Nudiviridae family, there are several well studied members that are capable of replicating in insect fat body, midgut and reproductive tissues and cause distinct pathologies.

Helicoverpa zea nudivirus–2 (HzNV-2) is a sexually transmitted nudivirus that causes anomalies in *H. zea* (corn earworm) reproductive organs. It also exhibits a biphasic

replication pattern where the infection is asymptomatic in latent phase and productive in lytic phase. Most HzNV-2 infections are asymptomatic having no morphological or physiological signs of pathology or detectable replication, however, almost one third of all HzNV-2 infections are productive (lytic) and this phase is characterized by fused gonadal and reproductive tract tissues that secrete virus particles in vesicles from the female genital opening that eventually forms a visible genital plug (Burand & Lu, 1997; Raina & Adams, 1995; Rallis & Burand, 2002b). Another symptom of infection is that the pheromone gland is hypertrophied, functional and produces elevated synthesis and release of pheromone. Females continue to call after mating attempts which enhances virus transmission to, and infection of, males which can then transmit virus to uninfected females in subsequent matings (Burand et al., 2005; Burand & Tan, 2006). Such changes in mating behavior, along with viral plug formation, help the virus to spread horizontally more efficiently. HzNV-2 is also transmitted vertically through generations via transovarial infections. In this case, the biphasic mode is dose-dependent and influenced by oviposition day (Burand & Rallis, 2004) with lower infectious doses and earlier oviposition days favoring latent infections. All these pathological features and host behavior modifications support rapid and efficient HzNV-2 transmission and persistence making it a potential candidate for development as a biopesticide for corn earworm management.

In a previous chapter, I used a more conventional PCR-based screening method to conduct a large-scale Heliothine nudivirus prevalence survey concentrated in the Cotton Belt but also including other regions of the U.S. I found that the HzNV-2 prevalence reaches up to 40% at some sampling locations in the cotton belt. I also detected significant nudivirus presence in a closely related host species, *Heliothis virescens* using conserved

nudivirus primers and subsequently HzNV specific PCR primers. In this chapter, we use bioinformatic tools and computational resources to survey for nudivirus presence in NCBI's lepidopteran sequence read archives (SRA) and nucleotide database. Additionally, we perform short nucleotide variation (SNV) analyses based on nudivirus sequences found in *H. armigera* and *H. zea* host populations to investigate differences in nucleotide polymorphisms between locations, host species and reference genomes. Finally, I consider possible hypotheses on nudivirus global distribution patterns inclusive of Asia, North and South America.

3.2 Material and Methods

In this "digital survey" of nudivirus species and distribution, I explored NCBI's (National Center for Biotechnology Information) two massive lepidopteran data sources: the Sequence Read Archive (SRA) and the nucleotide (NUC) database. These databases differ significantly in both data types and sources; the SRA database is exclusively for storing raw or minimally-processed public NGS data and the nucleotide database contains various relatively short individual sequences such as genes, short genomes, transcript data and sequences from different sources and studies. I developed two custom BASH (GNU Shell Emulator and Command Language Interpreter) scripts for data mining nudivirus sequences in these data sources. The first script (SRA_pipeline) executes a four-step workflow for each SRA experiment across the entire lepidopteran sequence database (Fig. 3.1). The second script (NUC_pipeline) follows a similar approach but uses different tools to collect all individual lepidopteran sequences into a large multi-FASTA file (Fig. 3.2). Following the pre-processing step, both scripts proceed with read-mapping in which short

nucleotide sequences are aligned to HzNV-1 and HzNV-2 reference genomes using the HiSAT2 program (D. Kim et al., 2015). This alignment step generates a sequence alignment/map (SAM) file that contains sequence information, quality scores, genomic coordinates and other details (H. Li et al., 2009) for each SRA experiment. Defining two reference genomes for sequence mapping, even though differences between them are subtle, remarkably increases the number of presumptive nudiviral sequence hits to the only the HzNV-1 genome but also generates many identical hits. In both pipelines, this problem was addressed by splitting the SAM files based on reference origin and these single-genome SAM files were then analyzed separately (Fig. 3.4).

3.2.1 SRA Database Mining

SRA mining was performed on NCBI's entire lepidopteran taxonomy with the help of a series of free and open source programs. A list of lepidopteran SRA experiments was downloaded through NCBI's web interface (https://www.ncbi.nlm.nih.gov/sra/?term=txid7088[Organism:exp) and these experiments were processed with a pipeline in which the output of a program acts as the input to another program. The SRA pipeline consists of four main steps; *i*) data download and conversion, *ii*) trimming and quality control, *iii*) alignment and homology checks, and iv) post-processing by removing tandem repeats, read duplicates and reference-based separation of final alignments. In the first step, raw SRA experiments are downloaded to LCC storage node and converted into FASTQ format using fastq-dump program (https://github.com/ncbi/sra-tools). The second step involves removal of sequencing adapters and low quality bases from the read ends using trim-galore software package

(Martin, 2011) with default quality scoring settings (Phred score: 20, error rate: 0.1, and a minimum length of 50 bp after trimming) (Fig. 3.3). In the third step, high quality SRA reads are mapped against both HzNV-1 and HzNV-2 genomes using a fast and sensitive short read aligner, HiSAT2 with relaxed mismatch penalty (default max. 6 and min 2; relaxed max. 2 and min. 1) and gap open/extend penalty (rdg and rfg; default value is 5, 3; relaxed value is 2,) parameters (Fig. 3.1). HiSAT2 is a splice-aware program so genomic and transcriptomic data can be processed concurrently. It also requires reference genomes to be sorted and indexed so another open-source program, samtools (H. Li et al., 2009) was used for these operations. Read alignments re-evaluated via blastn program (Camacho et al., 2009) with local HzNV-1 and HzNV-2 references and sequence homology was assessed based on alignment scores and e-value ($< e^{-10}$). Lastly, SAM files were post-processed using the GNU sed program (https://www.gnu.org/software/sed/) to remove uninformative tandem repeats (Fig 3.3) and 'rmdup' function in samtools to remove read duplicates (Fig. 3.4). Duplicate reads may occur in two ways; as a result of excessive PCR amplification and/or due to optical errors. PCR amplification is an essential step in some library preparation protocols. If two or more PCR amplimers attach to different spots on a flow cell, identical reads are generated during sequencing. Therefore, any duplicate variation in the original sequence will be overrepresented in the raw dataset. Additionally, optical duplicates may occur if the florescence from a single reaction spot is picked up by two or more sensors. These processes create two or multiple read duplicates which can be automatically detected and removed by samtools program. For reasons discussed above, the duplicate detection and removal step is crucial for all NGS datasets in order to minmize biases in downstream analysis.

Tandem repeats are a major issue in read mapping as they align to reference genome with high scores but a global BLAST query of these repeats usually yield no results or they align with many other distant organisms. These repeats were considered to be uninformative and were removed from the datasets. In addition to tandem repeat removal, paired-end reads were also decoupled because the paired-end reads actually contain two distinct sequences under a shared read code. This process reduces the complexity in the datasets so all the paired short reads treated as single reads after post-processing step. Moreover, SAM files were split into two based on reference origin via BASH script that reads "REF" information in each line and extracts that line to a separate file (Fig. 3.4). Also, I calculated the number of reads that uniquely matches to either HzNV-1 or HznV-2 by subtracting the number shared reads from all matches (n{HzNV1} + n{HzNV2} – n{HzNV1 ∩ HzNV2}) so multiple reads are avoided in combined analyses. Finally, after the read-mapping and post-process steps, the pipeline proceeds with the next SRA experiment in the list and runs the programs again with this new dataset. The alignment summary of all experiments was compiled using the 'stats' function which is a part of Samtools program.

3.2.2 Nucleotide Data Mining

Another custom BASH script (NUC_pipeline) was written for mining HzNV sequences from NCBI's Nucleotide database. Unlike SRA experiments, entries in this section contain a single fragment that ranges between several base pairs to large genomic assemblies of several megabases. In order to reduce the workload, sequences smaller than 1000 bp were not included in the mining process because these short nucleotide sequences

usually originate from specific studies which use specific PCR primers to amplify target regions so, technically, these amplimers do not contain any extraneous or endogenous viral sequence.

I followed a different approach for obtaining and preparing the nucleotide dataset (Fig. 3.2). First, I downloaded the entire lepidopteran sequence data (NCBI:txid7088) that is larger than 1000 bp and saved into a large multi-fasta file. This task was achieved via esearch and efetch utilities provided by the NCBI. Next, sequences in the multi-fasta file was split into 80 bp k-mers and labeled individually using pyfasta package (https://github.com/brentp/pyfasta). In general, all the deposited sequences in this section are well processed and curated so additional trimming and quality checks were not necessary. At this point, the structure of the dataset resembles a large NGS read file where each "short read" is originally a fragment of a lepidopteran DNA sequence. This batch processing approach was significantly faster than downloading and processing individual sequences due to the sheer number of submissions. Lastly, the multi-fasta file was converted to a large FASTQ file using seqtk program (https://github.com/lh3/seqtk) and the default quality scores were added to fulfill the FASTQ format requirements. Similar to SRA workflow, HiSAT2 program was used for reference mapping with relaxed setting and blastn software was employed for homology scoring (e-value < e^{-10}). The resulting SAM file contained sequences homologous to nudivirus DNA as well as uninformative tandem repeats which were removed by using GNU sed program (Fig. 3.3).

3.2.3 Variation Analysis

The aim of this procedure is to identify short nucleotide variations (SNVs) and predict the translational outcome of these variations by assessing the read alignment scores. This analysis procedure has three steps; reference-based separation of reads, and variant call, variant annotation and prediction. First, short nucleotide variations in the processed viral read datasets are evaluated via 'mpileup' function in samtools program with polyploidy parameter set to 1 (haploid mode). The 'mpileup' function distinguishes low quality polymorphisms from real variations by evaluating the alignment scores and read depth. This process generates a VCF (variant call format) which stores the coordinates and descriptions of nucleotide variations (Fig. 3.4).

Next, we used a separate program, SnpEff (Cingolani et al., 2012) to annotate and predict the effects of each variation for both HzNV-1 and HzNV-2 reference genomes (GenBank Accession: AF451898.1 and JN418988.1, respectively). SnpEFF uses genomic annotations to determine the location and the effect of nucleotide variation and then generates detailed reports about SNV summary, genomic coordinates, and changes in protein coding. These genomic annotations (GFF) were downloaded directly from NCBI's website and manually added to snpEff configuration file. SNP summary and statistics were obtained from each VCF file to compare the influence of host species, reference type and location.

3.2.4 Phylogeographic Analysis

Following the variation alanylsis, large viral datasets from Brazil were futher analyzed to investigate the viral phylogeny in conjunction with *Helicoverpa* hybridization

dynamics. Among these, a subset of 14 most infected datasets were selected based on viral read count (N > 1,000 reads) and total coverage (cov > 3). Since these libraries were generated based on specific restriction sites, the read coverage was biased and fragmented (Fig. 3.6). To overcome this issue, a consensus sequence for each SRA experiment in the subset was generated by using sam2consensus software (https://github.com/edgardomortiz/sam2consensus) while marking regions with "?" as missing data if there is no coverage. Phylogenetic analyses were done using BEAST, BEAUti and TreeAnnotator (Suchard et al., 2018) with total chain length of 1.1 M (10% burn-in rate), general time reversible substitution model and inverse gamma rate variation parameters. The geographical information, sampling locations and hybridization schemes were obtained from the original study (Cordeiro et al., 2020) and correlated with bioinformatic analyses.

3.2.5 Statistical Analysis

Statistical tests were performed using SNP percentages which were calculated by proportioning the number of SNPs to total dataset coverage. I used t-test and random effects ANOVA tests to compare sampling locations, host types and reference types based on SNP percentages. In order to compare host species, I used a t-test on *H. zea* ~ *H. armigera* dataset while excluding other parameters. A one-way ANOVA test was done using the entire data generated after "global" variant analysis with single independent variable and four groups; Global_Hzea_NV1, Global_Harm_NV1, Global_Harm_NV2, and Global_Harm_NV2. These results were further evaluated using a post hoc test using Tukey HSD method. The "global" data was partitioned into two independent variables (Location

and Host Species) and the reference genome was included as the third factor to perform a three-way ANOVA test on global dataset.

3.3 Results

3.3.1 SRA Dataset

The lepidopteran subcategory of NCBI's SRA database contains 35,039 independent submissions of NGS based experiments, as of October 2021. Among all SRA experiments analyzed, 2664 of them matched to HzNV genomic sequence. After removing the uninformative tandem repeats, we found 342,187 short reads in 694 different SRA datasets that are homologous to either HzNV-1 or HzNV-2 genome. The majority of these datasets were generated by three distinct research groups from China (NCBI: PRJNA730914, May 2021), Australia (Pearce et al., 2017) and Brazil (Gonçalves et al., 2019). The Brazilian study includes biosamples of *Helicoverpa zea* and *Helicoverpa armigera* and hybrids thereof. Besides these two major crop pests, there was evidence of nudiviruses in five other species; *Heliothis virescens*, *Helicoverpa assulta*, *Spodoptera frugiperda, Ostrinia nubilalis* and *Bombyx mori* (Table 3.1 and Table 3.2). Geographical distribution of the samples containing nudiviruses range from Brazil to China and Australia, Greece to the United States. The largest dataset in this survey was generated by researchers from Nanjing Agricultural University which contains 279 individual nudivirus infected experiments from *H. armigera* samples. Another large dataset was generated by researchers from University of Sao Paulo – ESALQ (BioProject: PRJNA615801) using both *H. armigera* and *H. zea* samples. This project contains 172 SRA experiments with 147 of them showing significant evidence of HzNV infections. Two of these experiments

by Cordeiro et al. (2020) SRR11432101 (*H. zea*) and SRR11432110 (*H. armigera*) were severely HzNV infected constituting 58.03% and 25.97% of all raw sequences respectively. Further variation analyses were performed on these two datasets (Table 3.4).

Helicoverpa zea. Among 118 SRA experiments analyzed, we found HzNV sequences in 107 of them. The mining procedure revealed 13341 unique short reads from HzNV origin with an average length of 212.2 bp. These experiments were conducted in many different locations including Brazil, Greece and the United States. Considering the individual SRA projects, the incident rates were 96.2% of projects in Brazil (51 out of 53 samples), 100% in Greece (8 out of 8) and 84.6% in the US locations (22 out of 26). Besides these major datasets, there are also several other experiments with relatively fewer read counts (Table 3.1 and Table 3.2).

Helicoverpa armigera. We analyzed 883 SRA experiments and found nudivirus presence in 497 of these experiments with total 98,593 unique reads with an average read length of 200.7 bp. Most of the infected *H. armigera* samples were originally from a single study conducted in China (Nanjing Agricultural University) and the incidence rate in this study was 94.7% (242 out of 249) with unique read counts ranging from 5 to 8371 per experiment. Additionally, 96 out of all 119 SRA experiments from Brazil exhibited HzNV sequences and in one experiment (SRR11432110), the number of unique read counts reached up to 6412 after the duplicate removal step. Similarly, several nudivirus sequences were detected in samples from Australia (Table 3.1)

Heliothis virescens. The number of SRA experiments deposited in NCBI's database was 31 with 19 of these showing nudivirus traces. The largest dataset in this category was originally published by researchers from Western Sydney University (NCBI:

PRJNA379496) and a total of 483 HzNV short reads were detected in 4 of their experiments. There are also traces of nudivirus sequences in several other *H. virescens* experiments from the US and Germany. Additionally, in this category, the number of the reads aligned to HzNV-1 genome was many times higher than the reads that aligned to HzNV-2 genome.

Bombyx mori. The domestic silkworm, *Bombyx mori*, is a well-studied beneficial insect that has more than 2200 SRA experiments deposited to the NCBI's database. Analysis of these experiments revealed many nudivirus sequences in 71 SRA experiments (3.16%). Most of the infected experiments were deposited by researchers from the Southwestern University (China) and included sequences from both BmE cell-lines and dissected silkworm body parts. Additionally, in a study based on BmE cell lines (NCBI: PRJNA518741) conducted by the Silkworm Genome Laboratory (China), 2 of the 10 deposited SRA experiments exhibited 9854 reads of presumptive nudivirus origin with over 92% query coverage and 95% sequence identity.

Spodoptera frugiperda. Bayer Crop Science deposited 30 *S. frugiperda* datasets to the NCBI database (BioProject: PRJNA545483). These datasets were generated from pooled samples (50 insects per sample) collected in Brazil. One dataset in this study (SRR9289276) yielded 6 short reads that are 150 bp long and covers 173 bp region, with more than 94% query coverage and 96% sequence identity to respective HzNV-2 genomic regions. These reads were aligned with ORF30 of HzNV-2 genome which is a hypothetical protein of unknown function.

Other groups. Several HzNV-like sequences detected in SRA experiments conducted with *Helicoverpa assulta* and *Ostrinia nubilalis* samples however these

sequences usually resemble conserved genes in their host genome (i.e. hydroxymethyltransferases) so they are considered as ambiguous or uninformative. Further analyses required for explaining the degree and origin of this type of sequence similarity.

3.3.2 SNP and Variant Analyses

Variant analysis of all 694 SRA experiments revealed 20,163 unique SNVs across the entire nudivirus reads. Among those, 3,335 SNVs were predicted as structural variants (Fig. 3.5). Comparison of nudivirus variations showed that *H. armigera* hosts have significantly higher SNP percentages than *H. zea* host species (t-test, df = 743, $p < 0.026$) however the result of the one-way ANOVA test showed significant differences in all pairwise comparisons except the "Global_Hzea_NV2-Global_Hzea_NV1" pair (df = 3, f value = 14.486, $p < 0.001$). On the other hand, the three-way ANOVA test showed that the sample location, reference genome and the interaction between them explains the variation in SNP percentages (Table 3.3). The interaction between location and reference genome indicates that the influence of location on SNP percentages is elevated with the addition of reference genome factor. We further analyzed two SRA experiments from Brazil that contained high numbers of nudivirus sequences (SRR11432101 and SRR11432110) and found that HzNV reads from *H. zea* sample showed a higher number of mutations than *H. armigera* samples in most common mutation types. In the high impact category which defines a group of variations that cause truncated protein or loss of function, *H. armigera* samples showed a higher mutation rate (Table 3.4). Additionally, no SNP effects were detected in smaller datasets due to low coverage rates, however, several high impact SNP variations that cause start/stop codon loss or frameshift variants detected in some *B. mori*

experiments (Southwest University – China and State Key Laboratory of Silkworm Genome Biology – China) and in some *H. armigera* datasets (BCM – Australia) (Table 3.5 and Table 3.6).

3.3.3 Phylogeographic Structure

The phylogeographic analysis was done based on the SRA data generated by Cordeiro et al., (2020) where these Brazilian researchers investigated hybridization and introgression dynamics between *H. armigera* and *H. zea.* The SRA experiments generated by these researchers contains large numbers of nudiviral sequences. In addition to their datasets, we followed their hybridization scheme to infer phylogeographic structure of HzNV strains. Figure 2 (Cordeiro et al., 2020) displays the high and low hybridization probability regions between *H. armigera* and *H. zea* populations. The cladogram and the hybridization probability map (Fig. 3.8) suggest a fairly low hybridization probability for SRR11432110.armigera (MTRL) branch and for most of the Piraricaba.zea (SPPI) clade. Additionally, higher levels of hybridization probabilities noted in the original study for branches and clades in-between. This information can also be interpreted as HzNV host range expansion due to hybridization and bilateral virus transmission.

3.3.4 Nucleotide Mining

As of October 2021, NCBI's lepidopteran nucleotide database contains 5,473,966 sequences with sizes range from 3 bp place holder master records to 128,845,201 bp chromosomal fragments. After removal of short sequences, the final dataset contained 1,122,994 distinct nucleotide sequences that are split into 80 bp fragments. Mapping these

fragments against HzNV-1 and HzNV-2 reference genomes revealed 200 homologous matches. After removing tandem repeats, the remaining 68 short sequences mapped to the HzNV genome with relatively high scores and low E-values (Table 3.7). Nine of these fragments showed homology to hydroxynethyltransferase enzymes which are also found in the host genome, and other sequences aligned to multiple different intergenic regions. Also, a majority of the short DNA sequences aligned to HzNV-1 genome (65 out of 68) with higher scores compared to a few aligned to HzNV-2 genome (3 out of 68) with high scores. Finally, these short sequences were from *Helicoverpa zea*, *Bombyx mori*, *Helicoverpa armigera*, *Heliothis virescens*, and *Operophtera brumata* (Table 3.7) host insects.

3.4 Discussion

Two large lepidopteran genetic sequence databases were investigated and a total of 342,187 unique nudivirus sequences were detected in 694 distinct SRA experiments. These sequences were mapped to either the HzNV-1 or HzNV-2 genomes with high alignment scores and considered as presence of genomic nudivirus-like sequences or in some severe cases an indication of an active nudivirus infection. Even though the high alignment scores and low BLAST e-values are determined for decision making process, false discovery rate adjustments were not applied to datasets due to sheer number of short reads and data complexity. Besides it's known host, *H. zea*, we found evidence for HzNV-like viruses infecting other host species notably *H. armigera*, *H. virescens*, and *B. mori* (insect and cell lines). We also found traces of nudivirus sequences in some "pooled" *Spodoptera frugiperda* samples that contained low numbers of nearly identical HzNV sequences.

Lastly we discarded samples from 18 other insect species (including *H. subflexa*, *H. assulta* and *B. mandarina*) due to low read numbers, poor alignment quality and thus lacking in clear evidence of nudivirus infection, although some potential nudivirus sequences were also detected in these insects.

We employed the two known HzNV strains to perform read-mapping in order to obtain as much nudivirus sequence as possible. Unexpectedly, when using the HzNV-1 genome as a reference, we detected almost 3 times more datasets than HzNV-2 reference alone. The number of the viral reads dropped by half when using HzNV-2 based analysis (Table 3.1 and Table 3.2). The HzNV-1 genome has two large deletions and several other smaller deletions compared to HzNV-2 genome but it is unlikely that these missing fragments are the main reason behind the difference in viral detection numbers. It is likely that the sequences found only in HzNV-1 are unrelated to the ancestral nudivirus genome but reflect sequences acquired during cell culture passages.

The digital survey of lepidopteran experiments generated large read datasets that allowed us to investigate nucleotide variations in multiple species collected from different locations. For each HzNV infected dataset, we generated a detailed report that describes the nucleotide variations and their effects on gene expression. Comparison of nucleotide variations in two largest datasets (*H. zea* and *H. armigera*) obtained from a study conducted in Brazil revealed that HzNV-2 found in *H. zea* host has more SNPs in almost all categories except "high impact" compared to *H. armigera* host. This pattern can be explained by the fact that *H. zea* is the native host for HzNV-2 and thus shows more sublethal variation relative to *H. armigera* which exhibits four high impact variations (i.e. loss of start/stop codon and/or frame-shift). Differences in nucleotide variation numbers are an indication of

founders effect occurred during bilateral transmission process. Additionally, we calculated SNP rates by dividing the number of total SNPs by the read coverage and we used it (in percentage form) as an indicator of SNP diversity in corresponding HzNV populations. One-factor ANOVA test of stacked SNP percentages showed significant difference between host species (Fig. 3.4). On the other hand, the results from the three factor ANOVA revealed that the SNP percentage is explained by location, reference and the interaction between them (Table 3.2). This result suggests that the variation among HzNV strains is primarily associated with the sampling location.

The datasets surveyed in this study were compiled from many distinct projects and experiments designed to answer different questions. As a result, most of the major data sources were fragmented both genetically (Figure 3.5 and Figure 3.6) and spatially (Fig. 3.7). For this reason, it was not possible to make sound inferences beyond nudivirus detection, however, we generated a cladogram of 14 high prevalence datasets using a Bayesian method that can appropriately process missing data (Fig. 3.8). This cladogram was congruent with the hybridization map provided by Cordeiro et al., 2020 and allowed us to predict potential interspecies transmission routes between feral *H. armigera* and *H. zea* populations in Brazil. Also the data suggest that bilateral virus transmission occurred via interspecies mating in high hybridization probability zones indicated in the original study.

Besides the experiments submitted to the SRA database, we also analyzed sequences in the nucleotide section by splitting each nucleotide sequence into 80 base k-mers. This method allowed us to process the entire set of lepidopteran sequences similar to SRA experiments and we found nudivirus-like sequences in *H. zea*, *H. armigera*, *H.*

virescens, *B. mori* and *O. brumata* specimens (Table 3.7). Despite the fact that a range of nucleotide submissions contain HzNV-like sequences, it was not entirely clear whether these sequences originate from the host or nudivirus, partly because a considerable knowledge gap exists in HzNV-2 gene functions and annotations.

In this chapter, we introduce a series of evidences showing that HzNV strains are circulating in many feral bollworm populations around the world (Table 3.1 and 3.2), and likely cause similar pathologies to those of HzNV-2 infections. The results presented in this study may be used as a starting point for investigating, isolating and propagating novel Heliothine nudiviruses. Also, the methods described here can be modified to analyze other pathogens in different arthropod groups based on a given viral or bacterial genome. Moreover, by incorporating a *de novo* analysis component, this method can be used to discover novel extrachromosomal genomes in datasets deposited to public databases.

In conclusion, I have developed a method for screening HzNV prevalence in public short read sequence (SRA) databases using a custom workflow based on existing bioinformatic tools and methods. I used read-mapping method to find viral sequences and BLAST tool to eliminate low score sequences from downstream analyses. This workflow can be adapted to other host and pathogen/parasitic taxa with only minor modification.

Figure 3.1: Script used for acquiring and analyzing the submissions in NCBI's SRA database

```
#!/bin/bash
## LCC SLRUM related code ###
## Read SRA code from the List

LIST=$1,
while IFS= read -r SRACODE
do

## Fetch Source (DNA/RNA) and Layout (Single/Paired) parameters
SOURCE=$ (curl -s https: //www.ncbi.nlm.nih.gov/sra/?term=$SRACODE | grep 'Source:' |
awk -F"<div>Source: <span>" '{print $2}' | awk -F"</span></div>" '{print $1}')
LAYOUT=$ (curl -s https: //www.ncbi.nlm.nih.gov/sra/?term=$SRACODE | grep 'Layout:' |
awk -F"<diveLayout: <span>" '{print $2}* | awk -F"</span></div>" '{print $1}')

## Load SRAtools and download the dataset
module load ccs/sratoolkit/sratoolkit.2.9.6-1
fasterq-dump $SRACODE -0 $SRACODE/sradump -t $SRACODE/tmp -e 8

## Load Trim-galore and trim based on Layout parameter
module unload ccs/sratoolkit/sratoolkit.2.9.6-1 & module load ccs/conda/trim-galore-0.6.4
if [ "$LAYOUT" = "SINGLE" ]; then
        echo "Layout: SINGLE"
        trim_ galore -j 8 --dont_gzip -o $SRACODE/trimmed $SRACODE/sradump/$SRACODE.fastq
        fastqc -t 8 $SRACODE/trimmed/*. fq
elif [ "$LAYOUT" = "PAIRED" ]; then
        trim_galore -j 8 --paired --dont_gzip -o $SRACODE/trimmed/ $SRACODE/sradump/* 1.
fastq. $SRACODE/sradump/*_2.fastq
        fastqc -t 8 $SRACODE/trimmed/*. tq
else
        echo "Trim QC error!!"

fi


## Load HiSAT2 and reference mapping
module unload ccs/conda/trim-galore-0.6.4 && module load ccs/conda/hisat2-2.1.0

if [ "$LAYOUT" = "SINGLE" ]; then
echo "Layout: SINGLE"
        hisat2 -p 8 --rdg 2,1 --rfg 2,1 --mp 1,0 --un $SRACODE/unmapped/unmap -x host_index/
hindex -U $SRACODE/trimmed/* trimmed. fq -S $SRACODE/tmp/map
        hisat2 -p 8 --rdg 2,1 --rfg 2,1 --mp 1,0 --al $SRACODE/virus_mapped/al_mapped -x
virus_index/vindex -U $SRACODE/unmapped/unmap* -S $SRACODE/virus_mapped/virus_mapped..sam
elif [ "SLAYOUT" = "PAIRED" ]; then
        hisat2 -p 8 --un-conc $SRACODE/unmapped/unmap -x host_index/hindex -1 $SRACODE/
trimmed/* 1 val_1.fq -2 $SRACODE/trimmed/* 2 val_2.fq -S $SRACODE/tmp/map
        hisat2 -p 8 --al-conc $SRACODE/virus_mapped/alconc_mapped -x virus_index/vindex -1
S$SRACODE/unmapped/*.1 -2 $SRACODE/unmapped/*.2 -S $SRACODE/virus_mapped/virus_mapped .sam
else
        'echo "Mapping error!!"
fi

## SAM to BAM conversion and sorting/indexing BAM files
module unload ccs/conda/hisat2-2.1.0 && module load ccs/conda/samtools-1.9
samtools view -@ 8 -S -b $SRACODE/virus_mapped/virus_mapped.sam > $SRACODE/
```

Figure 3.1 (continued)

```
virus_mapped/virus_mapped.bam
samtools sort -@ 8 $SRACODE/virus_mapped/virus_mapped.bam -o $SRACODE/virus_mapped/
virus_mapped_sorted.bam
samtools view -@ 8 $SRACODE/virus_mapped/virus_mapped_sorted.bam | awk '$2 == 0 || $2
== 16' >> results

## SAM to Fasta conversion for Blast search
awk '{print ">" $1 "\n" $10}' $SRACODE/virus_mapped/results.sam >> $SRACODE/
virus_mapped/results.fa

## Switch modules
module unload ccs/conda/samtools-1.9 && module load ccs/conda/blast-2.9.0
blastn -task blastn -query $SRACODE/virus_mapped/results.fa -db blastDB/vindex -
evalue 1e-20 -num_threads 16 -max_target_seqs 1 -outfmt 6 -out $SRACODE/virus_mapped/
blast_results.csv

## Remove duplicates (find unique IDs) from blast outfile
sort -u -t$'\t' -k1,1 $SRACODE/virus_mapped/blast_results.csv > $SRACODE/virus_mapped/
unique_blasts.csv

## Extract unique IDs colums from CSV file
cat $SRACODE/virus_mapped/unique_blasts.csv | cut -f1 -s > $SRACODE/virus_mapped/
uniqueIDcol.txt

## Switch modules
module unload ccs/conda/blast-2.9.0 && module load ccs/conda/samtools-1.9

## Copy SAM header and find reads from SAM file using unique IDs
head -n 3 $SRACODE/virus_mapped/virus_mapped.sam >> $SRACODE/virus_mapped/final.sam
samtools view $SRACODE/virus_mapped/virus_mapped.sam | grep -w -F -f $SRACODE/
virus_mapped/uniqueIDcol.txt >> $SRACODE/virus_mapped/final.sam
cp $SRACODE/virus_mapped/final.sam results/"$SRACODE"_"$SOURCE"_final.sam

## Cleaning (to avoid storage limit errors)
cp $SRACODE/trimmed/*_val_1_fastqc.html results/qc/"$SRACODE"_1_report.html
cp $SRACODE/trimmed/*_val_2_fastqc.html results/qc/"$SRACODE"_2_report.html
rm -rf $SRACODE/sradump
rm -rf $SRACODE/trimmed
rm -rf $SRACODE/unmapped
rm -rf $SRACODE/tmp
rm ~/ncbi/public/sra/$SRACODE.sra.cache

## Remove the SRA code from the list
sed -i /$SRACODE/d  $SRABATCH

done < $LIST
```

Figure 3.2: Script used for removing tandem repeats and short sequences

```
#!/bin/bash
## LCC SLRUM related code ###

## Read SRA code from the list

LIST=$1

while IFS= read -r SRACODE
do

sed -i '/TACCTACCTACCTACC\|TAGGTAGGTAGGTAGG\|TAAATATTAAATATTAAATATTAAATAT\|
ACTTACTTACTTACTT\|GTTGTTGTTGTTGTT/d'  data/$SRACODE

sed -i '/TAATAATAATAATAATAA\|TTGATGGATTGATGGATTGATGGA\|CAACAACACCAACAACAC\|
TCAAAATCAAAATCAAAATCAAAA/d'  data/$SRACODE

sed -i '/TACTACTACTACTAC\|TTAAATATTAAATATTAAATA\|TACACTACACTACAC\|GATTTTGATTTTGATTTT/
d'  data/$SRACODE

sed -i '/CAACAACACCAACAACAT\|CAACAACACCAC\|TAGTGGTAGTGGTAGTGG\|CATCATCATCATCAT\|
CAACAACAACAACAA/d'  data/$SRACODE

sed -i '/GATGATGACGATGATGAC\|TAATAATAATAATAA\|TTGTACTTGCACTTGTACTTGCAC\|
TATAAAATATAAAATATAAAA/d'  data/$SRACODE

sed -i '/GTATGTATGTATGTAT\|TAGTAGTAGTAGTAG\|TTTGGGTTTGGGTTTGGG\|CAACACCAACAACACCAA/
d'  data/$SRACODE

sed -i '/GAGGTTGAGGTTGAGGTT\|CCCAAACCCAAACCCAAA\|TAAGTAAGTAAGTAAG\|GATGATGATGAT\|
TTGTACTTGTACTTGTAC/d'  data/$SRACODE

sed -i '/TGAGGATGAGGATGAGGA\|TACCTACCTAAC\|TGTAATTGTAATTGTAAT\|TAGATAGATAGATAGA\|
TACCTACCTACC\|TAGTAGCATTAGTAGCAT\|TTGCACTTGCACTTGCACTTGCAC/d'  data/$SRACODE

sed -i '/AS:i:18\|/AS:i:19\|/AS:i:20\|/AS:i:21\|/AS:i:22\|/AS:i:23\|/AS:i:24\|/AS:i:
25\|/AS:i:26\|/AS:i:27\|/AS:i:28\|/AS:i:29\|/AS:i:30\|/AS:i:31\|/AS:i:32\|/AS:i:33\|/
AS:i:34\|/AS:i:35\|/AS:i:36\|/AS:i:37\|/AS:i:38\|/AS:i:39\|/AS:i:40\|AS:i:41\|AS:i:
42\|AS:i:43\|AS:i:44\|AS:i:45\|AS:i:46\|AS:i:47\|AS:i:48\|AS:i:49\|AS:i:50/d'
data/$SRACODE



done < $LIST
```

Figure 3.3: Script used for duplicate removal and variant call

```
#!/bin/bash
## LCC SLRUM related code ###
## Read SRA code from the list

LIST=$1
while IFS= read -r SRACODE
do

## Copy dataset to temp folder and split based on reference genome
cp  data/$SRACODE.sam temp/"$SRACODE"_nv1.sam
cp  data/$SRACODE.sam temp/"$SRACODE"_nv2.sam
sed -i '/JN418988\|@PG/d' temp/"$SRACODE"_nv1.sam
sed -i  '/AF451898\|@PG/d' temp/"$SRACODE"_nv2.sam

## Remove duplicates and variant call using 'mpileup' function on HzNV-1 alignments
samtools rmdup -S temp/"$SRACODE"_nv1.sam temp/"$SRACODE"_nv1_nodup.sam
samtools view -S -b temp/"$SRACODE"_nv1_nodup.sam > temp/"$SRACODE"_nv1_nodup.bam
samtools sort temp/"$SRACODE"_nv1_nodup.bam -o temp/"$SRACODE"_nv1_nodup_sorted.bam
samtools index temp/"$SRACODE"_nv1_nodup_sorted.bam
cp temp/"$SRACODE"_nv1_nodup_sorted.bam  results/nv1/"$SRACODE"_nv1_nodup_sorted.bam
samtools view results/nv1/"$SRACODE"_nv1_nodup_sorted.bam > results/
nv1/"$SRACODE"_final.sam
rm results/nv1/"$SRACODE"_nv1_nodup_sorted.bam
bcftools mpileup -Ou -f refs/nv1.fasta temp/"$SRACODE"_nv1_nodup_sorted.bam |
bcftools call --ploidy 1 -mv -Ob -o temp/"$SRACODE"_nv1_calls.bcf
bcftools convert -O v -o temp/"$SRACODE"_nv1_calls.vcf temp/"$SRACODE"_nv1_calls.bcf
sed -i 's/AF451898/AF451898.1/g' temp/"$SRACODE"_nv1_calls.vcf
cd snpEff && java -jar snpEff.jar -v AF451898 -stats ../results/
html/"$SRACODE"_nv1_annotated.html -csvStats ../results/
csv/"$SRACODE"_nv1_annotated.csv ../temp/"$SRACODE"_nv1_calls.vcf > ../results/nv1/
nv1_vcf/"$SRACODE"_nv1_annotated.vcf && cd ..

## Remove duplicates and variant call using 'mpileup' function on HzNV-2 alignments
samtools rmdup -S temp/"$SRACODE"_nv2.sam temp/"$SRACODE"_nv2_nodup.sam
samtools view -S -b temp/"$SRACODE"_nv2_nodup.sam > temp/"$SRACODE"_nv2_nodup.bam
samtools sort temp/"$SRACODE"_nv2_nodup.bam -o temp/"$SRACODE"_nv2_nodup_sorted.bam
samtools index temp/"$SRACODE"_nv2_nodup_sorted.bam
cp temp/"$SRACODE"_nv2_nodup_sorted.bam  results/nv2/"$SRACODE"_nv2_nodup_sorted.bam
samtools view results/nv2/"$SRACODE"_nv2_nodup_sorted.bam > results/
nv2/"$SRACODE"_final.sam
rm results/nv2/"$SRACODE"_nv2_nodup_sorted.bam
bcftools mpileup -Ou -f refs/nv2.fasta temp/"$SRACODE"_nv2_nodup_sorted.bam |
bcftools call --ploidy 1 -mv -Ob -o temp/"$SRACODE"_nv2_calls.bcf
bcftools convert -O v -o temp/"$SRACODE"_nv2_calls.vcf temp/"$SRACODE"_nv2_calls.bcf
sed -i 's/JN418988/JN418988.1/g' temp/"$SRACODE"_nv2_calls.vcf
cd snpEff && java -jar snpEff.jar -v JN418988 -stats ../results/
html/"$SRACODE"_nv2_annotated.html -csvStats ../results/
csv/"$SRACODE"_nv2_annotated.csv  ../temp/"$SRACODE"_nv2_calls.vcf > ../results/nv2/
nv2_vcf/"$SRACODE"_nv2_annotated.vcf && cd ..

## Clean-up temp folder (not included)
## Extract SNP and coverage information
samtools sort results/nv1/nv1_sam/$SRACODE.sam -o temp/$SRACODE.bam
RESULT1=$(samtools coverage -H temp/"$SRACODE"_nv1.bam)
echo "$SRACODE, $RESULT1" >> coverage_nv1.csv

samtools sort results/nv2/nv2_sam/$SRACODE.sam -o temp/$SRACODE.bam
RESULT2=$(samtools coverage -H temp/"$SRACODE"_nv2.bam)

echo "$SRACODE, $RESULT2" >> coverage_nv2.csv
done < $LIST
```

Table 3.1: Summary of nudivirus reads that align to HzNV-1 genome

Scientific Name	Center Name	Total Datasets	Total Reads	Mean Read Length (bp)	Mean Genome Coverage (%)
Bombyx mori	NCBI GEO	9	94	144.56	0.05
	Jiangsu University of Science And Technology	10	173	246.40	0.05
	Shenyang Agricultural University	1	2	299.00	0.05
	Southwest University	56	53381	245.71	6.19
	State Key Laboratory of Silkworm Genome Biology	2	4432	280.50	10.53
	UT-GALS	2	35	200.00	0.05
Helicoverpa armigera	BCM (Australia)	28	849	278.86	0.08
	Central China Normal University	1	5	90.00	0.11
	China Agricultural University	5	15	69.60	0.05
	Chines Academy of Agricultural Sciences	3	27	300.00	0.09
	Chinese Academy of Agricultral Sciences	30	471	300.80	0.09
	CSIR-National Chemical Laboratory(CSIR-NCL)	11	468	202.00	0.09
	CSIRO	9	90	199.89	0.09
	Embrapa Genetic Resources And Biotechnology	6	237	200.00	0.09
	NCBI GEO	10	270	231.00	0.09
	Hainan University	18	323	299.00	0.09
	Henan Agricultural University	5	75	100.00	0.08
	Institute of Plant Protection, Chinese Academy of	1	23	180.00	0.09
	Institute of Zoology, CAS	16	69	170.13	0.07
	Nanjing Agricultural University	242	83414	321.79	1.76
	Sun Yat-Sen University	4	22	200.50	0.10
	University of Sao Paulo-ESALQ	84	7185	168.95	1.88

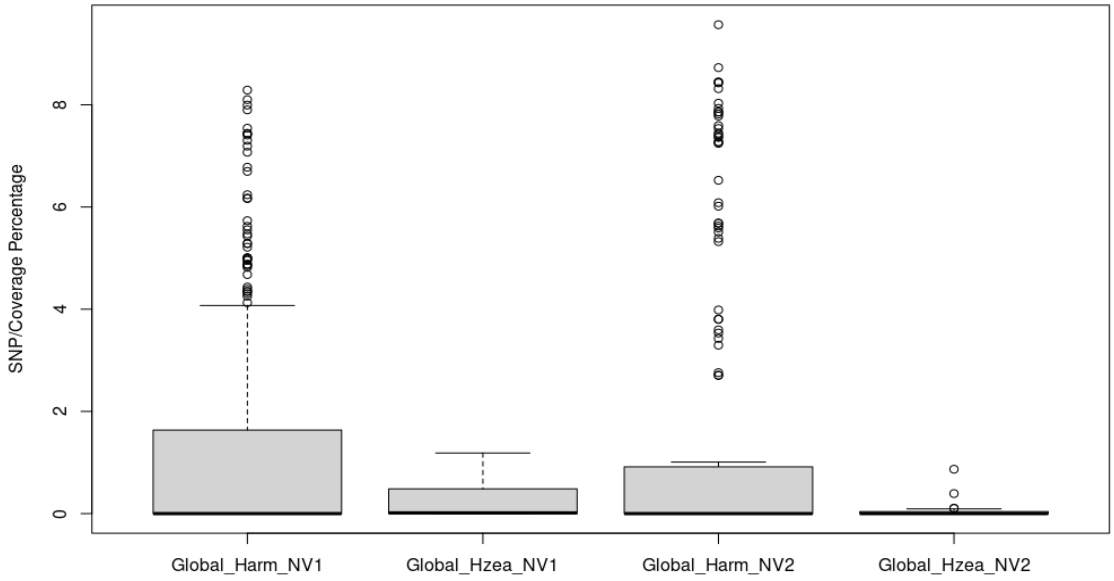
Table 3.1 (continued)

Helicoverpa assulta	CSIRO	4	31	197.50	0.07
	Institute of Plant Protection, Chinese Academy of	2	10	200.00	0.08
	Max-Planck-Institute For Chemical Ecology, Germany	1	1	202.00	0.04
Helicoverpa zea	Agricultural Research Service, US Department of Ag	4	20	222.25	0.10
	BCM (Australia)	7	2767	426.29	0.20
	Foundation For Research And Technology Hellas (Greece)	8	1735	202.00	0.26
	Iowa State University	9	573	302.00	0.25
	The University of Queensland	1	288	200.00	0.26
	University of Maryland	9	574	302.00	0.21
	University of Sao Paulo-ESALQ	51	15479	168.82	7.05
	USDA-ARS	18	3462	84.28	0.24
Heliothis virescens	American Museum of Natural History	3	127	250.00	0.15
	NCBI GEO	1	2	151.00	0.13
	Max-Planck-Institute For Chemical Ecology, Germany	6	48	201.00	0.10
	Max-Planck-Institute For Chemical Ecology, Germany	4	32	198.00	0.09
	University of Maryland, College Park	1	32	202.00	0.05
	Western Sydney University	4	483	1161.75	0.13
Spodoptera frugiperda	Bayer Crop Science	1	2	150.00	0.05
Total Result		694	177326	255.81	1.93

Table 3.2: Summary of nudivirus reads that align to HzNV-2 genome

Scientific Name	Center Name	Total Datasets	Total Reads	Mean Read Length (bp)	Mean Genome Coverage (%)
Bombyx mori	NCBI GEO	8	22	150.13	0.05
	Jiangsu University of Science And Technology	4	21	202.00	0.05
	Southwest University	38	67144	223.55	7.13
	State Key Laboratory of Silkworm Genome Biology	2	8854	280.50	11.88
	UT-GALS	2	6	200.00	0.05
Helicoverpa armigera	BCM (Australia)	21	43364	577.29	8.29
	CSIR-National Chemical Laboratory (CSIR-NCL)	2	2	202.00	0.04
	NCBI GEO	1	1	200.00	0.04
	Institute of Plant Protection, Chinese Academy Of	1	1	180.00	0.04
	Institute of Zoology, CAS	1	1	194.00	0.03
	Nanjing Agricultural University	73	151998	300.00	6.43
	University of Sao Paulo-ESALQ	68	12605	168.76	3.89
Helicoverpa zea	University of Sao Paulo-ESALQ	52	27828	168.83	11.79
Heliothis virescens	Max-Planck-Institute for Chemical Ecology, Germany	8	17	200.50	0.07
	University Of Maryland, College Park	1	13	202.00	0.05
Ostrinia nubilalis	Tufts University	3	5	166.67	0.04
Spodoptera frugiperda	Bayer Crop Science	2	5	150.00	0.06
Total Result		287	311885	242.57	6.33

Figure 3.4: Comparison of SNP percentages of all viral reads using ANOVA test with single independent variable with four groups.



*Legend: Global_[Host Species]_[Reference Genome]

A) Analysis of Variance Table
Response: values

	Df	Sum Sq	Mean Sq	F valuePr(>F)	
ind	3	165.25	55.084	14.486	3.369e-09
Residuals	809	3076.26	3.803		

B) Tukey multiple comparisons of means
(95% family-wise confidence level)
Fit: aov(formula = values ~ ind, data = global.stacked)

	diff	lwr	upr	p adj
Global_Hzea_NV1-Global_Harm_NV1	-0.8138783	-1.3498560	-0.2779007	0.0005802
Global_Harm_NV2-Global_Harm_NV1	0.5241246	0.0739552	0.9742941	0.0148735
Global_Hzea_NV2-Global_Harm_NV1	-0.9976383	-1.7300210	-0.2652555	0.0026909
Global_Harm_NV2-Global_Hzea_NV1	1.3380030	0.7163686	1.9596374	0.0000002
Global_Hzea_NV2-Global_Hzea_NV1	-0.1837599	-1.0323830	0.6648632	0.9445142
Global_Hzea_NV2-Global_Harm_NV2	-1.5217629	-2.3189716	-0.7245542	0.0000064

Table 3.3: Summary of the three-factor ANOVA test

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Location	4	398.7	99.66	29.558	< 2e-16
HostSpecies	1	0.1	0.13	0.039	0.843
Reference	1	67.8	67.76	20.096	8.45e-06
Location:Reference	2	84.0	41.99	12.452	4.73e-06
HostSpecies:Reference	1	0.2	0.25	0.074	0.786
Residuals	796	2684.0	3.37		

SNP percentage dataset was analyzed using three-factor ANOVA test and found significant difference between locations, reference genomes and the interaction between these two factors. The interaction between location and reference genome indicates that the influence of location on SNP percentages is elevated with the addition of reference genome factor.

Table 3.4: SNP annotations and effect predictions of two largest HzNV datasets from Brazil based on HzNV-2 reference genome.

Host: *H. zea* (SRR11432101)

Number of effects by type and region

	Count	*%*
conservative_inframe_deletion	1	0
conservative_inframe_insertion	1	0
disruptive_inframe_deletion	2	0
disruptive_inframe_insertion	1	0
downstream_gene_variant	3186	0.45
frameshift_variant	1	0
intergenic_region	519	0.07
missense_variant	289	0.04
splice_region_variant	4	0
stop_retained_variant	4	0
synonymous_variant	362	0.05
upstream_gene_variant	2760	0.39

Number of effects by impact

	Count	*%*
HIGH	1	0
LOW	366	0.05
MODERATE	294	0.04
MODIFIER	6465	0.91

Number of effects by functional class

	Count	*%*
MISSENSE	289	0.44
SILENT	366	0.56

Host: *H. armigera* (SRR11432110)

Number of effects by type and region

	Count	*%*
conservative_inframe_deletion	1	0
conservative_inframe_insertion	1	0
disruptive_inframe_deletion	2	0
downstream_gene_variant	2948	0.43
frameshift_variant	2	0
intergenic_region	579	0.08
missense_variant	226	0.03
splice_region_variant	4	0
start_lost	1	0
stop_lost	1	0
stop_retained_variant	3	0
synonymous_variant	300	0.04
upstream_gene_variant	2756	0.4

Number of effects by impact

	Count	*%*
HIGH	4	0
LOW	303	0.04
MODERATE	230	0.03
MODIFIER	6283	0.92

Number of effects by functional class

	Count	*%*
MISSENSE	228	42.938%
SILENT	303	57.062%

Figure 3.5: Alignment overview of two large datasets from Brazil to HzNV-1 genome.

A) Nudivirus sequences from H. armigera (SRR11432110) host before duplicate removal

B) Nudivirus sequences from H. armigera (SRR11432110) host after duplicate removal

C) Nudivirus sequences from *H. zea* (SRR11432101) host before duplicate removal

D) Nudivirus sequences from *H. zea* (SRR11432101) host after duplicate removal

Figure 3.6: Alignment overview of two large datasets from Brazil to HzNV-2 genome.

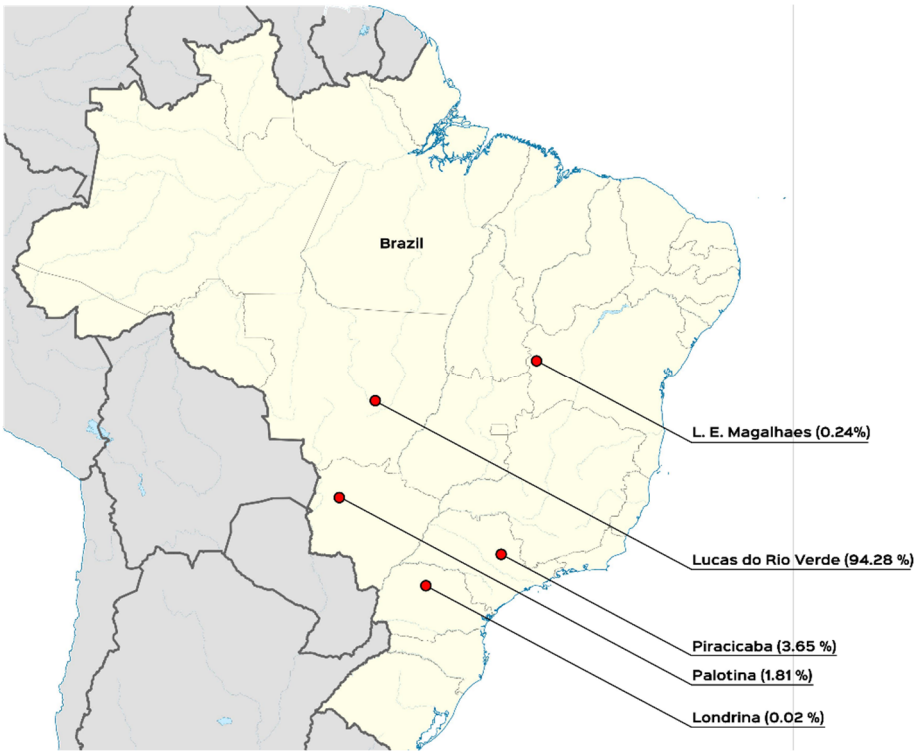
A) Nudivirus sequences from H. armigera (SRR11432110) host before duplicate removal

B) Nudivirus sequences from H. armigera (SRR11432110) host after duplicate removal

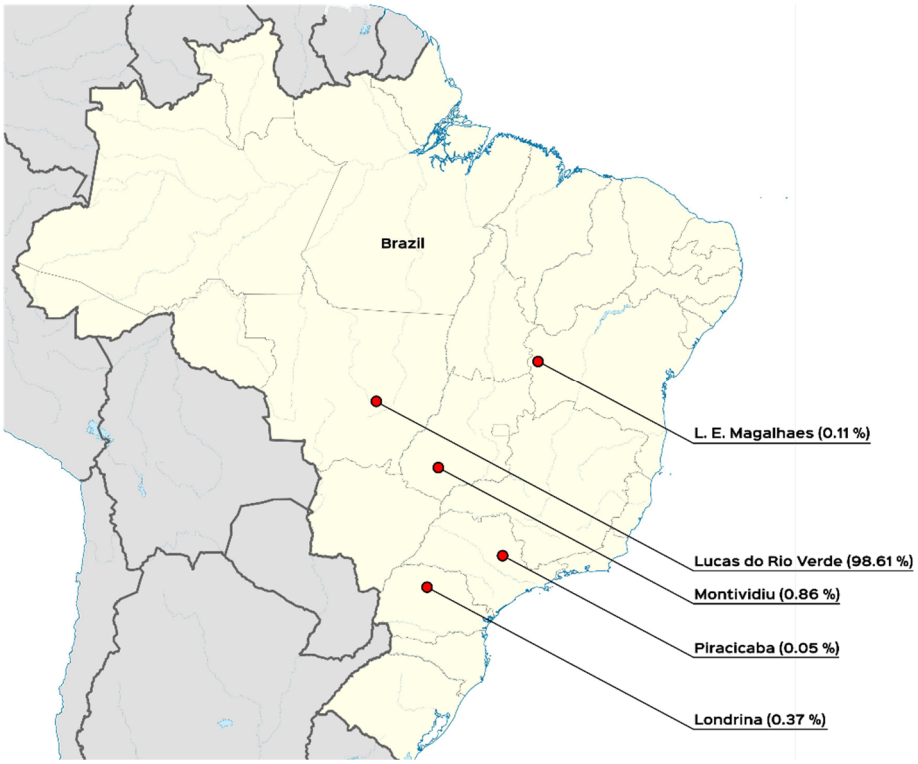
C) Nudivirus sequences from *H. zea* (SRR11432101) host before duplicate removal

D) Nudivirus sequences from *H. zea* (SRR11432101) host after duplicate removal

Figure 3.7: Sampling locations and nudivirus prevalence for samples collected from Brazil.

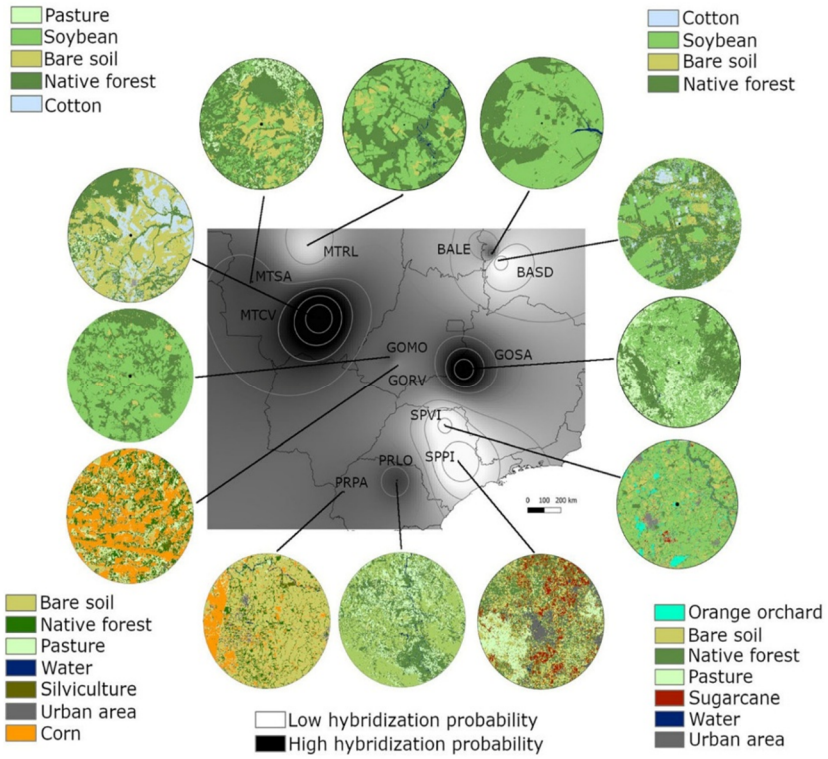


A) *H. zea*



B) *H. armigera*

Figure 3.8: Sampling locations with hybridization probabilities (modified from Cordeiro et al., 2020) along with the cladogram generated with 14 largest HzNV datasets. Lighter shades indicate lower, darker shades indicate higher hybridization probability.



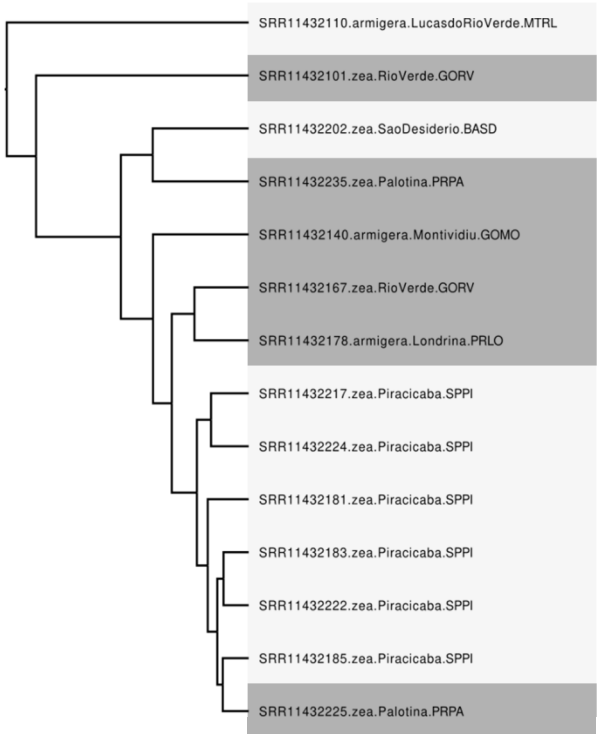


Table 3.5: Summary of variant analysis results (HzNV-1 reference only)

Scientific Name	Center Name	SNP Effects (Avg.) HIGH	MED	LOW	Exp. Count	Read Count
Bombyx mori	NCBI GEO	0.00	2.70	2.80	10	94
	Jiangsu University of Science And Technology	0.00	5.70	7.30	10	173
	Shenyang Agricultural University	0.00	0.00	0.00	1	2
	Southwest University	1.68	135.19	450.19	57	53381
	State Key Laboratory of Silkworm Genome Biology	4.50	397.50	1379.5	2	4432
	UT-GALS	0.00	9.00	9.00	2	35
Helicoverpa armigera	BCM (Australia)	0.00	0.00	0.00	28	849
	Central China Normal University	0.00	1.00	9.00	1	5
	China Agricultural University	0.00	0.00	0.00	5	15
	Chinese Academy of Agricultural Sciences	0.00	0.00	0.00	33	498
	CSIR-National Chemical Laboratory (CSIR-NCL)	0.00	0.00	0.00	11	468
	CSIRO	0.00	0.00	0.00	9	90
	Embrapa Genetic Resources and Biotechnology	0.00	0.00	0.00	6	237
	NCBI GEO	0.00	0.00	0.00	10	270
	Hainan University	0.00	0.00	0.00	18	323
	Henan Agricultural University	0.00	0.00	0.00	5	75
	Institute of Plant Protection, CAS	0.00	0.00	0.00	1	23
	Institute of Zoology, CAS	0.00	0.00	0.00	16	69
	Nanjing Agricultural University	0.39	45.56	168.29	249	83464
	Sun Yat-Sen University	0.00	0.00	0.00	4	22
	University of Sao Paulo-ESALQ	0.09	2.39	2.52	92	7185
Helicoverpa zea	Agricultural Research Service, US Department of Ag	0.00	0.00	0.00	4	20
	BCM (Australia)	0.86	0.29	0.00	7	2767
	Foundation For Research And Technology Hellas	0.75	0.63	0.00	8	1735
	Iowa State University	0.33	0.22	0.00	9	573
	The University of Queensland	0.00	0.00	0.00	1	288
	University of Maryland	0.33	0.44	0.00	9	574
	University of Sao Paulo-ESALQ	0.28	6.08	6.77	53	15479
	USDA-ARS	0.22	0.28	0.00	18	3462
Heliothis virescens	American Museum of Natural History	0.00	0.00	0.00	3	127
	NCBI GEO	0.00	0.00	0.00	1	2
	Max-Planck-Institute For Chemical Ecology, Germany	0.00	0.6	1.50	11	80
	University of Maryland, College Park	0.00	4.00	9.00	1	32
	Western Sydney University	0.00	0.00	0.00	4	483
Total Result		0.35	29.35	101.66	699	177332

Table 3.6: Summary of variant analysis results (HzNV-2 reference only)

Scientific Name	Center Name	SNP Effects (Avg.) HIGH	MED	LOW	Exp. Count	Read Count
Bombyx mori	NCBI GEO	0.00	0.70	1.00	10	22
	Jiangsu University of Science and Technology	0.00	0.00	0.00	9	21
	Shenyang Agricultural University	0.00	0.00	0.00	1	0
	Southwest University	2.37	147.50	472.63	38	67144
	State Key Laboratory of Silkworm Genome Biology	7.00	612.00	1920.0	2	8854
	UT-GALS	0.00	0.00	0.00	2	6
Helicoverpa armigera	BCM (Australia)	1.13	79.27	264.03	63	43364
	Central China Normal University	0.00	0.00	0.00	1	0
	China Agricultural University	0.00	0.00	0.00	5	0
	Chinese Academy of Agricultral Sciences	0.00	0.00	0.00	30	0
	CSIR-National Chemical Laboratory(CSIR-NCL)	0.00	0.00	0.00	11	2
	CSIRO	0.00	0.00	0.00	9	0
	Embrapa Genetic Resources and Biotechnology	0.00	0.00	0.00	6	0
	NCBI GEO	0.00	0.00	0.00	10	1
	Hainan University	0.00	0.00	0.00	18	0
	Institute of Plant Protection, CAS	0.00	0.00	0.00	1	1
	Institute of Zoology, CAS	0.00	0.00	0.00	15	1
	Nanjing Agricultural University	0.49	58.29	207.10	249	151998
	Sun Yat-Sen University	0.00	0.00	0.00	4	0
	University of Sao Paulo-ESALQ	0.05	2.83	3.54	92	12605
Helicoverpa zea	Agricultural Research Service, US Department of Ag	0.00	0.00	0.00	4	0
	BCM (Australia)	0.00	0.00	0.00	13	0
	Foundation For Research and Technology Hellas	0.00	0.00	0.00	8	0
	Iowa State University	0.00	0.00	0.00	9	0
	The University of Queensland	0.00	0.00	0.00	1	0
	University of Maryland	0.00	0.00	0.00	9	0
	University of Sao Paulo-ESALQ	0.08	8.92	11.26	53	27828
	USDA-ARS	0.00	0.00	0.00	18	0
Heliothis virescens	NCBI GEO	0.00	0.00	0.00	1	0
	Max-Planck-Institute For Chemical Ecology, Germany	0.00	0.00	0.00	11	17
	University of Maryland, College Park	0.00	4.00	11.00	1	13
	Western Sydney University	0.00	0.00	0.00	4	0
Total Result		0.43	38.25	128.45	708	311877

Figure 3.9: Script for acquiring and analyzing the submissions in NCBI's nucleotide database

```
#!/bin/bash

## LCC SLURM related codde

module load ccs/conda/samtools-1.9

ACCBATCH=$1

while read -r ACCODE
do

A=$(efetch -db nuccore -id $ACCODE -format fasta)
B=$(sed -e "1q"  <<< "$A" | sed -e 's/[\t ]//g;/^$/d')
C=$(sed '3,$s/^/'"$B"'_linenum\n/' <<< "$A" | sed '1 s/^.*$/'$B'/' | awk
'{gsub("linenum",NR,$0);print}')

echo "$C" > $ACCODE.fasta

magicblast -num_threads 2 -no_unaligned -gapopen 0 -gapextend 0 -penalty -1 -query
$ACCODE.fasta -db blastDB2/vindex2 -out "$ACCODE"_results.sam

rm -f $ACCODE.fasta

countt=$(samtools view -c "$ACCODE"_results.sam)

if [ $countt == 0 ]
then
 rm -f "$ACCODE"_results.sam
else
 mv "$ACCODE"_results.sam results/"$ACCODE".sam
fi

sed -i /$ACCODE/d  $ACCBATCH

done < $ACCBATCH
```

Table 3.7: BLAST search results of nucleotide database mining study. Query matches were identified based on available gene annotations. *A: glycine hydroxymethyltransferase, B: serine hydroxymethyltransferase

	Reference	Host	Seq. length	Percent Identity	E-value	Reference Gene	Gene Func.*	NCBI Code
1	HzNV-1	*H. zea*	77	84.416	3.11E-13	ORF72	A	KZ118848.1
2	HzNV-1	*O. brumata*	76	85.526	8.91E-14	ORF72	A	JTDY01008121.1
3	HzNV-1	*O. brumata*	63	90.476	8.91E-14	ORF72	A	JTDY01003972.1
4	HzNV-1	*O. brumata*	63	87.302	4.62E-11	ORF72	A	JTDY01008121.1
5	HzNV-1	*B. mori*	75	82.667	1.61E-10	ORF72	A	DF090336.1
6	HzNV-1	*B. mori*	77	83.117	1.32E-11	ORF72	A	ICPK01019056.1
7	HzNV-2	*B. mori*	77	81.818	1.61E-10	ORF66	B	ICPK01019056.1
8	HzNV-2	*B. mori*	75	81.333	1.96E-9	ORF66	B	DF090336.1
9	HzNV-1	*O. brumata*	63	88.889	3.79E-12	intergenic		JTDY01003972.1
10	HzNV-2	*O. brumata*	63	85.714	1.96E-9	ORF66	B	JTDY01008121.1
11	HzNV-1	*O. brumata*	76	84.211	1.09E-12	intergenic		JTDY01008121.1
12	HzNV-1	*H. zea*	77	83.117	1.32E-11	intergenic		KZ118848.1
13	HzNV-1	*H. zea*	55	94.545	3.11E-13	intergenic		KZ118493.1
14	HzNV-1	*H. zea*	48	100	8.91E-14	intergenic		KZ118493.1
15	HzNV-1	*H. armigera*	81	91.358	2.73E-20	intergenic		KZ149999.1
16	HzNV-1	*H. armigera*	81	91.358	2.73E-20	intergenic		NW_018395591.1
17	HzNV-1	*H. zea*	88	89.773	4.32E-24	intergenic		KZ118158.1
18	HzNV-1	*H. zea*	80	100	3.79E-31	intergenic		KZ118493.1
19	HzNV-1	*H. zea*	80	100	3.79E-31	intergenic		KZ118493.1
20	HzNV-1	*H. zea*	51	100	2.1E-15	intergenic		KZ118139.1
21	HzNV-1	*H. zea*	55	98.182	6.01E-16	intergenic		KZ118283.1
22	HzNV-1	*H. zea*	66	100	1.51E-23	intergenic		KZ116754.1
23	HzNV-1	*H. zea*	72	98.611	3.55E-25	intergenic		MT702890.1
24	HzNV-1	*H. zea*	72	100	8.34E-27	intergenic		KZ117299.1
25	HzNV-1	*H. zea*	73	100	2.39E-27	intergenic		KZ118535.1
26	HzNV-1	*H. zea*	73	100	2.39E-27	intergenic		MT702914.1
27	HzNV-1	*H. zea*	80	100	3.79E-31	intergenic		KZ116722.1
28	HzNV-1	*H. zea*	80	98.75	1.61E-29	intergenic		KZ118283.1
29	HzNV-1	*H. zea*	80	93.75	1.51E-23	intergenic		KZ118158.1
30	HzNV-1	*H. zea*	80	98.75	1.61E-29	intergenic		KZ118139.1
31	HzNV-1	*H. zea*	72	97.222	4.32E-24	intergenic		KZ118493.1
32	HzNV-1	*H. zea*	72	94.444	2.24E-21	intergenic		KZ118493.1

Table 3.7 (continued)

33	HzNV-1	*H. zea*	77	100	1.61E-29	intergenic	MT702890.1
34	HzNV-1	*H. zea*	80	98.75	1.61E-29	intergenic	KZ118283.1
35	HzNV-1	*H. zea*	80	98.75	1.61E-29	intergenic	KZ116754.1
36	HzNV-1	*H. zea*	80	97.5	1.96E-28	intergenic	MT702890.1
37	HzNV-1	*H. zea*	80	97.5	1.96E-28	intergenic	KZ117299.1
38	HzNV-1	*H. zea*	80	98.75	1.61E-29	intergenic	MT702914.1
39	HzNV-1	*H. zea*	82	95.122	1.02E-25	intergenic	KZ118535.1
40	HzNV-1	*H. zea*	80	97.5	1.96E-28	intergenic	KZ116722.1
41	HzNV-1	*H. zea*	80	93.75	5.26E-23	intergenic	KZ118283.1
42	HzNV-1	*H. zea*	80	93.75	4.32E-24	intergenic	KZ116739.1
43	HzNV-1	*H. zea*	80	85	8.91E-14	intergenic	KZ117675.1
44	HzNV-1	*H. zea*	73	98.63	1.02E-25	intergenic	KZ118139.1
45	HzNV-1	*H. zea*	58	100	3.32E-19	intergenic	KZ116754.1
46	HzNV-1	*H. zea*	54	100	4.93E-17	intergenic	MT702890.1
47	HzNV-1	*H. zea*	51	100	2.1E-15	intergenic	KZ117299.1
48	HzNV-1	*H. zea*	49	100	2.55E-14	intergenic	KZ118535.1
49	HzNV-1	*H. zea*	68	98.529	6.41E-22	intergenic	KZ118283.1
50	HzNV-1	*H. zea*	71	97.183	1.84E-22	intergenic	MT702890.1
51	HzNV-1	*H. virescens*	72	100	8.34E-27	intergenic	NWSH01002613.1
52	HzNV-1	*H. zea*	80	100	3.79E-31	intergenic	KZ118251.1
53	HzNV-1	*H. virescens*	80	98.75	1.61E-29	intergenic	NWSH01002613.1
54	HzNV-1	*H. virescens*	80	98.75	1.61E-29	intergenic	NWSH01002551.1
55	HzNV-1	*H. virescens*	65	95.385	4.05E-18	intergenic	NWSH01001560.1
56	HzNV-1	*H. zea*	73	100	2.39E-27	intergenic	KZ117975.1
57	HzNV-1	*H. zea*	80	100	3.79E-31	intergenic	KZ118251.1
58	HzNV-1	*H. virescens*	80	97.5	1.96E-28	intergenic	NWSH01002551.1
59	HzNV-1	*H. virescens*	80	93.75	5.26E-23	intergenic	NWSH01000291.1
60	HzNV-1	*H. virescens*	80	98.75	1.61E-29	intergenic	NWSH01001560.1
61	HzNV-1	*H. virescens*	75	92	3.32E-19	intergenic	NWSH01000053.1
62	HzNV-1	*H. zea*	61	100	7.81E-21	intergenic	KZ117975.1
63	HzNV-1	*H. virescens*	80	93.75	4.32E-24	intergenic	NWSH01001560.1
64	HzNV-1	*H. zea*	80	100	3.79E-31	intergenic	KZ118251.1
65	HzNV-1	*H. virescens*	80	95	1.02E-25	intergenic	NWSH01001560.1
66	HzNV-1	*H. zea*	80	100	3.79E-31	intergenic	KZ118251.1
67	HzNV-1	*H. virescens*	87	89.655	5.26E-23	intergenic	NWSH01001560.1
68	HzNV-1	*H. virescens*	67	98.507	1.84E-22	intergenic	NWSH01001560.1

CHAPTER 4: SEQUENCE CHARACTERIZATION AND ANALYSIS OF THE HELIOTHIS VIRESCENS NUDIVIRUS

4.1 Introduction

Corn earworm (*Helicoverpa zea*) is a polyphagous Noctuid species that poses a serious threat to many economically important cash crops such as corn, cotton, tobacco, and soybean. The estimated damage caused by corn earworm infestations and cost of control in the United States alone exceeds $2 billion per year (Pogue, 2004). In early instars, the corn earworm larva enters the fruiting body of plants and continuously feeds on vegetative and reproductive tissues until pupation. During this feeding stage, corn earworm larvae are often protected from environmental and behavioral factors which render many conventional management methods infective. In addition to management issues caused by its feeding behavior, *H. zea* populations can develop "field evolved" metabolic resistance against multiple Bt toxins and conventional insecticides fairly quickly which further complicates their management strategies. Recent advancements in Bt technologies and new "Bt-pyramid" products can provide significant protection against corn earworm damage, however selection pressure caused by excessive use of these products is likely to result in practical resistance in the near future.

Even though insects are known to transmit human, animal and plant diseases, they are also susceptible to many viral and bacterial pathogens. There are more than 1100 known species of insect viruses that infect more than 20 insect families (Grzywacz, 2017). Almost half of these viruses belong to Baculoviridae family which is a group of viruses known for their polyhedral occlusion bodies. These occlusion bodies provide a protective layer against environmental degradation and improves *per os* infectivity. Baculoviruses, along

with Bracoviruses, diverged from an ancestral Nudivirus strain around 280 Mya (Thézé et al., 2011). The Nudiviridae family contains several non-occluded and facultatively-occluded viruses that cause a range of pathological symptoms during their infections. The first known nudivirus, *Oryctes rhinoceros* (OrNV) was discovered in 1966 during an extensive survey of coconut palm beetle and its pathology was described shortly after in severely lethargic larvae with disintegrated fat body tissues (Huger, 2005). *Oryctes rhinoceros* (coconut palm beetle) causes devastating losses in palm plantations in the Middle East and South East Asia. Pilot studies conducted in Samoan islands showed the potential of OrNV in coconut palm beetle management by reducing the palm tree damage by 95% in some cases (Bedford, 1980).

The first sequenced nudivirus, HzNV-1 was discovered in an established lepidopteran cell line and identified as a persistent infectious agent that episodically exhibits lytic infections similar to Baculoviruses in cell culture. The HzNV-1 genome closely resembles the HzNV-2 but HzNV-1 can only replicate in some insect cell lines such as IMC-Hz and TN-368 (Lin et al., 1999). The HzNV-1 genome has two large deletions in addition to several short nucleotide variations compared to HzNV-2 genome and that may have been generated during cell immortalization (IMC-Hz). The genome sequence and potential reading frames of HzNV-1 were identified using a targeted shotgun-based sequencing method (Cheng et al., 2002).

The *Helicoverpa zea* nudivirus (HzNV-2) is a sexually transmitted double-strand DNA virus that can sterilize its host by infecting gonad tissue and causing tissue destruction, fusions and anomalies in the reproductive tracts. The HzNV-2 also modifies reproductive behavior in female corn earworm moths by promoting excessive pheromone

synthesis (Burand et al., 2005; Rallis & Burand, 2002b) and by blocking the production and function of pheromonostatic peptide (Burand & Tan, 2006; Kingan et al., 1995). Additionally, HzNV-2 exhibits a biphasic replication cycle where it may cause lytic infections that lead to sterile agonadal phenotypes or alternately produce latent phase infections that are asymptomatic without detectable morphological symptoms.

The two HzNV strains were sequenced in the early 2000's using shotgun-sequencing methodologies that consisted of random shearing of the viral genome, size fractionation, cloning and sequencing random clones with the resulting sequences then assembled to produce a first draft of the genome sequence (Cheng et al., 2002; Burand et al., 2012). These steps are required for target enrichment that is followed by dye-termination sequencing. In contrast, modern flow cell based NGS technologies can generate massive amounts of sequence reads randomly from a library prepared in a relatively quick and affordable way. NGS libraries can be prepared to sequence whole genomes or specific regions of DNA or RNA sources. Specific regions can be enriched via PCR-based or hybridization-based methods to obtain significantly more read data from those regions. On the other hand, whole genome sequencing approaches require random fragmentation of all genetic material in the sample which produces concurrent sequence from any extrachromosomal DNA in the sample.

The word "meta" is a Greek prefix which means "after" or "beyond" and metagenomics is the term that has been coined to describe genomic approaches for studying the multiple genomes simultaneously and sequences produced by NGS sequencing and extends the focus of sequencing projects "beyond" the target species. Metagenomic approaches are especially important for studying microbial communities and infectious

diseases that otherwise require more intensive time and resource consuming procedures. Metagenomics also provides tools and protocols for studying population genetics of self-replicating microbes, viruses, and quasispecies models and even defective interfering particles found in many viral infections (Alnaji et al., 2019; Vignuzzi & López, 2019).

In previous chapters, we present evidence that HzNV-like viruses circulate in *Helicoverpa armigera, Heliothis virescens* and *Bombyx mori* populations in Brazil, China and Australia. Some *Helicoverpa armigera* samples (i.e. SRR11432110) were dominated by HzNV sequences, however, it was impossible to recover a complete *H. armigera* nudivirus genome from those datasets due to target enrichment methods used in library preparations. The "original" study that produced these datasets were designed to investigate inter-species hybridization patterns by a genotyping-by-sequencing (GBS) method which employs a PCR-based enrichment step to generate more read depth around specified endonuclease restriction sites. However, GBS approaches can also create significant gaps in genomic assembly.

In this chapter, we present the complete genomic sequence of a novel nudivirus, the *Heliothis virescens* nudivirus (HvNV) that infects *H. virescens* populations in the southeast United States. We use a set of *de novo* assemblers with metagenomic parameters to analyze the NGS data we generated from a randomly fragmented nudivirus-positive DNA library. Our results show that the HvNV genome is similar to the HzNV-2 genome in terms of gene ontology and sequence identity, however short nucleotide variations are common when the two genomes are compared. Additionally, our *de novo* assembly workflow revealed another small circular sequence similar to HvNV genome which resembles a defective interfering particle with several missing open reading frames. Based on genomic structure

and sequence similarity, it can be predicted that HvNV will exhibit similar pathological symptoms to the HzNV-2 virus as genes known or suspected to be involved in host pathology are present in both viruses. If this prediction is correct, the pathology of HvNV may also be modified or engineered to make them a viable, host-specific and effectively spreading tobacco budworm management agent.

4.2 Materials and Methods

4.2.1 Sample Collection and Virus Detection

The first batch of *Heliothis virescens* samples used for sequencing were collected from Louisiana State University Red River Research Station (Bossier City) using bollworm traps that were baited with *H. virescens* specific pheromone lures. These pheromone traps were checked at least once every three days and trap contents transferred into a plastic container and stored in a -20°C freezer until shipping. These samples were exposed to environmental factors and showed DNA degradation at different levels so I decided to collect live insects to obtain higher quality DNA and potentially infectious virus particles. This second batch of samples were collected in August 2020 by Mr. Bentley Fitzpatrick from Louisiana State University Macon Ridge Research Station (Winnsboro, LA) and the insects were immediately frozen to minimize DNA degradation. This second batch was also shipped in a dry ice filled insulated box via overnight service.

Upon receiving the samples, abdomens were separated from thorax while frozen and transferred to individual 1.5 mL microfuge tubes. This procedure was performed in our laboratory using sterilized dissection instruments that were cleaned and heat-sterilized prior to and after each dissection. The abdomens were then homogenized in 200 µL sterile

phosphate-buffered saline (PBS) solution using pipette tips. PBS has previously been a widely used for both isolating and storing live nudivirus as well as extracting the DNA that is required for downstream applications. For DNA extraction, 50 µL of abdominal homogenate was incubated in an alkaline digestion solution (Table 2.1) at 56°C for 2 hours. Following the incubation period, total DNA was isolated as described in Table 2.1

Molecular screening to identify samples containing virus was performed by multiplex PCR using P4 and P13 primers (Table 2.2) which target conserved HzNV genes and generate 404 bp (ORF78) and 320 bp (ORF16) amplimers, respectively. This multiplexed approach was preferred as it reduced potential false negatives. In order to confirm host species identity, another chain reaction was performed with nudivirus positive samples using *H. virescens* specific primers (Table 2.2) which distinguishes *H. zea* (448 bp) from *H. virescens* (413 bp) based on a 35 bp deletion in *H. virescens* genome.

4.2.2 DNA Extraction and Sequencing

Genomic DNA was collected from *H. virescens* tissue homogenates using a spin-column based extraction method (DNeasy Blood & Tissue Kit, QIAGEN) that provides clean and pure DNA without use of phenol or chloroform. In this method, insect sample (tissue homogenate) is digested with a lysis buffer that contains detergents, salts, metal chelators, and proteinase K. This step is necessary to disrupt cellular structure, release the DNA and inhibit nuclease activity. After 2 hours of incubation at 56°C, the digested solution is transferred to a spin-column and the total DNA was captured on a silica membrane. Binding is maintained in the presence of chaotropic guanidine salts which create a molecular affinity of the DNA molecules for the membrane. While the DNA is

captured by the membrane, cellular debris and other contaminants are removed with washing solutions provided in the kit. The captured DNA is then released by removing salt ions from the solution which renders the silica membrane negatively charged.

The quality and purity of the DNA isolation was assessed using a spectrophotometer (Nanodrop 2000c, Thermo Scientific) and the DNA integrity was visualized by gel electrophoresis. The isolates were then submitted to a sequencing facility (Novogene, UC Davis) to complete the library preparation step and sequencing procedures. The Illumina libraries were prepared from randomly fragmented DNA and short read datasets were generated on NovaSeq (Illumina) platform in 150 bp size with paired-end setting (150 bp x 2, forward and reverse).

4.2.3 Data Analysis

Upon data acquisition, raw reads and sequencing reports were directly downloaded to LCC computer cluster in FASTQ format using wget utility. Before further analysis, datasets were preprocessed by removing platform specific adapters (Table 4.1) and low quality reads. This task was performed by trim-galore software on both read pairs with quality threshold value (PHRED) set to 20. The Trim-galore package also generates a post-trimming quality report that contains multiple statistics, sequencing scores and metrics about data quality and integrity. After the pre-processing step, paired short reads in FASTQ files were ready for haplotype calling and phylogenetic analysis.

HISAT2 is a fast and efficient read aligner (Kim et al., 2015) that generates SAM alignment files from read datasets and reference genomes. We used the HiSAT2 program to quantify the amount of nudivirus sequences in FASTQ files by aligning short reads

against HzNV-2 reference genome. Next, we examined viral haplotypes in the datasets using SPAdes; a *de novo* assembly program that performs well with metagenomic datasets (Sutton et al., 2019). This assembly method is fairly slow and uses significantly more computer resources compared to reference-based assemblers which can be biased towards a reference genome (Deng et al., 2021; Eliseev et al., 2020). The SPAdes program was used (Prjibelski et al., 2020) for de novo assembly of metagenomic reads. We then used another *de novo* assembler, MEGAHIT (Li et al., 2015) to compare and contrast the results. These two assemblers utilize unique graph-based algorithms (metaSPAdes and succinct *de Bruijn* graph, respectively) and they can reconstruct haplotypes from metagenomic datasets (Sutton et al., 2019). In general, SPAdes generates fewer contigs but consumes more computer power. On the other hand, MEGAHIT performs faster and uses less resources but generates excessive numbers of smaller contigs.

In SPAdes analysis, we used "--isolate" and "--metaviral" parameters to filter out chromosomal sequences and run the assembly pipeline using the remaining circular and extrachromosomal DNA (Antipov et al., 2020). Similarly, the MEGAHIT program was used with "–meta-sensitive" preset (kmer sizes 21,31,41,51,61,71,81,91,99) to recover viral haplotypes. In order to reduce computational load, circular genomes were linearized based on the synteny block structure of the HzNV-2 genome. This linearization step was performed by Mauve (Darling et al., 2004) genome aligner and the final alignment visualized by wgVista (Dubchak et al., 2009) whole genome alignment program.

4.2.4 Phylogenetic Analysis

In a previous chapter, we found significant amounts of nudivirus sequences in multiple SRA experiments submitted by several research groups globally. Consensus sequences were generated from those nudivirus infected datasets and they were aligned against the HzNV-2 genome using Geneious Prime 2021.2 (Kearse et al., 2012) program. Since the SRA datasets analyzed in the previous chapter were highly fragmented, only a few regions in HzNV-2 genome yielded the multiple sequence alignments necessary for phylogenetic analysis. The largest alignment block was selected for the analysis and a phylogenetic tree was generated using PHYML program (Guindon et al., 2010) which uses a maximum likelihood algorithm to calculate phylogenetic trees. No outgroup was defined for this analysis and the substitution model was set to Generalized Time Reversible (GTR). The final phylogenetic tree was visualized via figtree program (https://github.com/rambaut/figtree).

4.3 Results

4.3.1 Summary of the Datasets

Two sets of samples were collected by Bentley Fitzpatrick between 2019 and 2020 from two Louisiana State University research stations. The first batch of samples consisted of 227 tobacco budworm (*H. virescens*) males collected from Red River Research Station (Bossier City, LA) in 2019. The second batch consisted of 79 live captures collected between early-August to mid-September from Louisiana State University Macon Ridge Research Station (Winnsboro, LA). A total of 306 trap-captured *H. virescens* samples were screened using HzNV specific primers and the prevalence of *H. virescens* infecting

nudivirus was 4.8% and 5.1% in 2019 and 2020 samples, respectively. The number of nudivirus positives in the first batch was 11 and 2 of those showed strong PCR bands upon gel electrophoresis. Similarly, three nudivirus positives were detected in the second batch of samples, however, none of these yielded strong positives or enough DNA for the whole genome sequencing. Two strong positives from the first batch (LA27 and LA73) were selected for downstream analyses even though gel electrophoresis indicated signs of some DNA degradation. Despite concerns about sample inegrity, sequencing libraries were prepared and submitted. The Illumina platform generated 64.4 million and 61.5 million short reads for LA27 and LA73 libraries, respectively. Given the read size (150 bp) and the host genome size (~350 Mbp), mean sequencing coverage was 27.6x for the LA27 and 26.3x for LA73 datasets. The amount of adapters and low quality bases in raw datasets was very low and the final size of the preprocessed datasets were 99.6% and 99.57% for LA27 and LA73, respectively compared to raw datasets (Table 4.1).

In the next step, viral sequences were quantified by mapping the preprocessed reads against the HzNV-2 genome. In the LA27 dataset, less than 0.001% of all reads (719 total) were of viral origin and the coverage for this dataset was only 10.9% (25,251 bp). This dataset was discarded due to insufficient data points. In contrast, nearly 0.004% of the LA73 library was viral reads (242,649 total), the coverage was 99.1% (231,768 bp), and the average read depth for this dataset was 153.23. The HzNV-2 genome was not fully represented by the short reads due to deletions and insertions found in HvNV genome relative to the reference genome. These complete and defective HvNV genome sequences have been deposited to Genbank under accession #2528081.

4.3.2 Viral Haplotype Analysis

Viral haplotypes were reconstructed using both SPAdes and MEGAHIT *de novo* assemblers with metagenomic parameters. SPAdes assembler yielded 53 distinct scaffolds ranging from 1,136 bp to 231,768 bp. Batch homology search of these sequences resulted in one main circular viral haplotype (231,768 bp) and two other smaller nudivirus assemblies (166,099 bp and 166,051 bp) (Table 4.2). The fourth largest scaffold was 15,402 bp in size and matched with several mitochondrial sequences in BLAST homology search with high scores. The remaining contigs were smaller than 10 Kbp and they did not show homology to HzNV-2 genes (Table 4.2). Similar to SPAdes results, MEGAHIT *de novo* assembler produced only two nudivirus haplotypes that are 231,236 bp and 165,444 bp long. The total number of contigs were 538 and the next largest fragment in this list was 43,093 bp long which was aligned to *Hydraecia micacea* chromozome Z with high homology score (total score: 19445, E value: 0). Even though MEGAHIT generated an excessive number of contigs, results from SPAdes and MEGAHIT analyses were parallel and they supported the presence of two nudiviral haplotypes in LA73 sample. Since the results were almost identical, we used the SPAdes outputs for further analysis.

The main HvNV haplotype was 147 bp larger than the HzNV-2 genome and the pairwise sequence identity was 93.52%. Gene structure and synteny were similar between HvNV and HzNV-2 (Fig. 4.2), however the sequence similarity was lower than 90% in 8 out of 113 open reading frames (ORF) (Table 4.3). Moreover, some start and stop codons in the HvNV genome differed from that of the HzNV-2 genome. These codon differences were in ORF6 (start codon), ORF32 (stop codon), ORF57 (start and stop codons), ORF58

(stop codon), ORF69 (stop codon), ORF90 (start codon), ORF92 (start and stop codons), ORF107 (stop codon), and ORF 112 (stop codon).

The SPAdes assembly resulted in two small nudivirus haplotypes (Table 4.2, #2 and #3) that were identical in terms of sequence similarity and structure and the only difference was the two short deletions at the ends of the assembly. Due to this similarity, the smaller assembly (Table 4.2, #3) was excluded from downstream analyses. The larger haplotype showed similar gene synteny with HzNV-2 (Fig. 4.3) but it was 65,628 bp shorter than HzNV-2 genome. Pairwise sequence similarity between HvNV and the defective genome was 31.4%. Also, this defective haplotype was missing 25 open reading frames and the remaining genes showed significant differences in start and stop codon coordinates (Table 4.4).

4.3.3 Phylogenetic Analysis

We generated a phylogenetic tree using the sequences from nudivirus infected SRA experiments in addition to HzNV-2 and HvNV sequences. As discussed before, the SRA datasets were too fragmented to perform a phylogenetic analysis, however, I found several regions with sequence data from multiple SRA experiments. Among these regions, ORF38 had enough sequence data and spatial diversity to perform phylogenetic analysis (Fig. 4.4). I built a multiple sequence alignment using 10 homologus sequences from those deposited by researchers from China, Brazil and the US. Next, I performed a phylogenetic analysis with this alignment using PhyML program, a maximum likelihood algorithm that performs well without any predefined outgroup sequence. In the resulting phylogenetic tree, nudiviruses from different hosts clustered together except for the samples from the

Americas, which indicates that interspecies transmission occurred more recently compared to other groups (Fig. 4.5).

4.4 Discussion

HzNV-2 is one of the few insect viruses that spreads sexually and causes reproductive sterility. It is the only known sexually transmitted viral pathogen among all Heliothine bollworm species. HvNV-2 modifies host behavior to improve horizontal transmission efficiency and is also one of the few insect viruses that exhibits a biphasic replication cycle with an asymptomatic latent phase and a reproductive lytic phase. HzNV-2 can be modified to elevate its sterilizing effect (Webb, personal communication) and may eventually be used as a corn earworm biopesticide. Even though the HzNV-2 pathology and transmission is well studied, there is basically no information about other nudiviruses that are circulating in lepidopteran groups.

In previous chapters we have shown that nudiviruses are present in many Lepidoptera species and populations of those species around the world. The Lepidoptera now known to have nudivirus sequences include *Helicoverpa armigera, Heliothis virescens, Bombyx mori* and likely *Spodoptera frugiperda.* It is certain, that this is an incomplete list that is biased by the species currently represented in lepidopteran genome databases. Additionally, I showed that HzNV-2 is potentially circulating in many *Helicoverpa zea* populations outside the North America such as Greece, Brazil, and Australia. I also found traces of nudivirus sequences in some "pooled" *Spodoptera frugiperda* samples which contains low numbers of highly identical HzNV sequences. In this chapter we sequenced a novel nudivirus strain that was found in *Heliothis virescens*

insects collected from Louisiana. Genomic structure and composition of this novel nudivirus, HvNV closely resembles the HzNV-2 genome with only 6.48% difference in sequence identity.

Even though HvNV and HzNV-2 look similar in terms of sequence identity, several open reading frames in HvNV genome are different in size mostly due to differences in start and stop codon locations. Among these coding sequences, ORF90 and ORF92 has significantly large deletions and low sequence similarity (79.37% and 77.99% respectively) compared to homologous genes in the HzNV-2 genome. These two medium-size reading frames are missing in HzNV-1 which is restricted to insect cell lines. Gene products of these ORFs are still unknown but they are most likely involved in virus-host interaction. Also, based on the evidence presented in this chapter, we can argue that HvNV and other Heliothine nudiviruses found in the Americas have evolved recently from an ancestral nudivirus strain compared to the nudiviruses found in the Old World. This claim is also congruent with continental isolation between two ecoregions, however, the data we used for phylogenetic analysis was fairly limited and may not reflect the complete evolutionary relationships between the nudiviruses found in the Old World and in the Americas.

In addition to the main haplotype, we found a smaller nudivirus haplotype that is structurally and functionally homologous to the HvNV genome. This small sub-viral genome appears to be from a defective interfering particle (DIP) which is defined as small noninfectious viral particles that require a homologous ‘helper’ virus to replicate but also interferes with the replication of its helper (Treuhaft & Beem, 1982). Baculovirus DIPs can interfere with homologous virus replication and cause significant declines in bioreactor

production. In addition to DIPs, there are many sub-viral particles such as defective viral genomes (DVGs) and Von Magnus particles that are generated during replication cycles. These particles can also play important roles in virus pathology ranging from persistence to virulence modulation and interference to interferon-induction (Vignuzzi & López, 2019). By definition, the defective haplotype we have found in *H. virescens* host falls into the defective viral genome (DVG) category and could play a significant role in nudivirus pathology. Further experiments and metagenomic sequence data is necessary to understand the interaction between nudivirus DVGs and their homologs.

Table 4.1: Summary of whole genome sequencing and quality control results

	Total Basepairs	Read Total	Avg. Read Size	Read Size Std	Av. Read Quality	Read Quality Std
LA27_F Raw	4835570100	32237134	150	0	35.6426	1.59538
LA27_F Trimmed	4814216745	32234810	149.348	2.86023	35.661	1.53188
LA27_R Raw	4835570100	32237134	150	0	35.4189	1.57409
LA27_R Trimmed	4818965059	32234810	149.496	2.46648	35.4351	1.51702
LA73_F Raw	4618224300	30788162	150	0	35.5809	1.58202
LA73_F Trimmed	4596857378	30784887	149.322	2.68479	35.5964	1.52781
LA73_R Raw	4618224300	30788162	150	0	35.3009	1.73809
LA73_R Trimmed	4600578797	30784887	149.443	2.61675	35.3183	1.67942

Illumina adapters used in this sequencing run:

5' Adapter (5’ to 3’):

AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGT

ATCATT;

3' Adapter (5’ to 3’):

GATCGGAAGAGCACACGTCTGAACTCCAGTCACGGATGACTATCTCGTATGC

CGTCTTCTGCTTG.

Table 4.2: List of all extrachromosomal contigs and their BLAST search results (LA73).

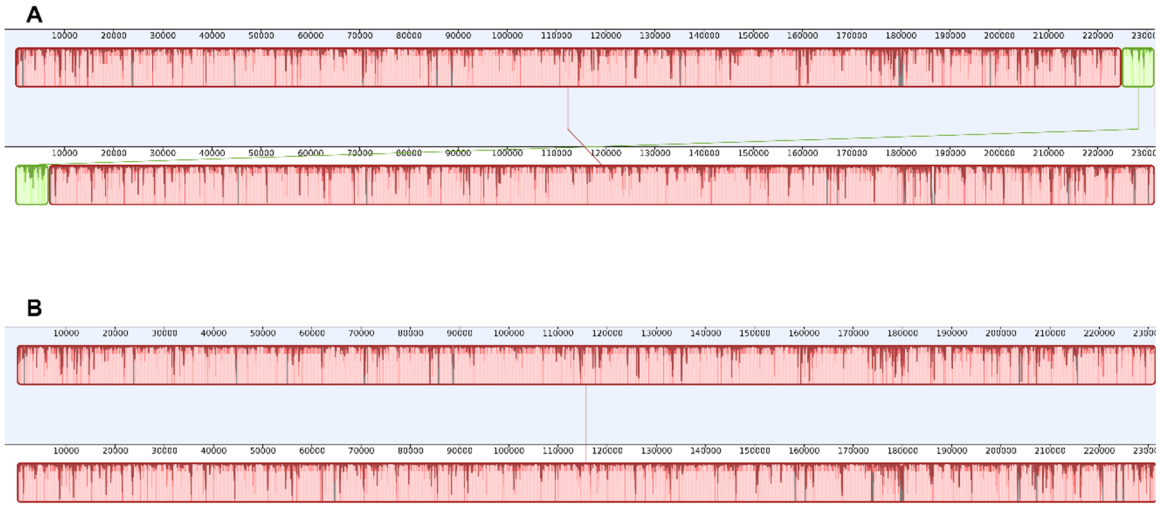
	Size (Bp)	COV	Cutoff	Type	Scientific Name	Total Score	Coverage	E value	Identity
1	231768	19.28	0	circular	*Helicoverpa zea nudivirus 2*	4.25E+05	99.00%	0	96.84%
2	166099	47.53	0	circular	*Heliothis zea virus 1*	36607	30.00%	0	82.36%
3	166051	47.53	33	circular	*Heliothis zea virus 1*	36607	30.00%	0	82.36%
4	15402	130.0	0	circular	*Mamestra brassicae*	95554	99.00%	0	89.58
5	8686	3.31	0	circular	*Trichoplusia ni*	4351	84.00%	0	77.67
6	8407	6.81	5	circular	*Hydraecia micacea*	10117	100.00%	0	87.72
7	7886	15.25	5	circular	*Hydraecia micacea*	11577	89.00%	0	79.75
8	7719	6.47	5	circular	*Noctua pronuba*	1.15E+05	98.00%	0	92.97
9	7347	3.26	0	circular	*Autographa pulchrina*	25944	99.00%	0	93.02
10	7037	7.98	5	circular	*Mythimna impura*	20342	88.00%	0	93.46
11	6888	23.54	5	circular	*Helicoverpa armigera*	3126	41.00%	0	86.51
12	6489	8.6	5	circular	*Abrostola tripartita*	8934	98.00%	0	90.86
13	6476	1.42	0	circular	*Mellicta athalia*	6192	80.00%	0	87.99
14	6299	1.72	5	circular	*Ochropleura plecta*	17710	99.00%	0	91.04
15	6155	1.47	5	circular	*Ochropleura plecta*	6405	90.00%	0	87.33
16	6066	1.41	0	circular	*Autographa pulchrina*	33149	94.00%	0	89.73
17	5988	16.09	5	circular	*Abrostola tripartita*	9815	100.00%	0	96.19
18	5934	1.46	0	circular	*Colias croceus*	8871	70.00%	0	87.4
19	5748	2.48	5	circular	*Hecatera dysodea*	5527	99.00%	0	81.14
20	5733	7.5	5	linear	*Schrankia costaestrigalis*	6753	77.00%	0	81.73
21	5664	1.03	0	circular	*Autographa pulchrina*	23860	99.00%	0	92.13
22	5575	7.31	5	circular	*Atethmia centrago*	5093	77.00%	0	88.15
23	5437	14.75	5	circular	*Autographa pulchrina*	7726	99.00%	0	92.99
24	5255	1.75	5	circular	*Mamestra brassicae*	5450	100.00%	0	85.46
25	5210	1.73	0	circular	*Lycaena phlaeas*	4501	89.00%	0	85.58
26	5180	31.48	5	circular	*Hypena proboscidalis*	6894	95.00%	0	91.82
27	5171	7.48	5	circular	*Eupsilia transversa*	9077	89.00%	0	82.75
28	5121	2.71	0	circular	*Autographa pulchrina*	6946	98.00%	0	91.08
29	4967	22.97	5	circular	*Autographa pulchrina*	457	12.00%	8.00E-67	85.38

Table 4.2 (continued)

30	4923	7.77	5	circular	*Idaea aversata*	5296	94.00%	0	88.07
31	4661	2.96	5	circular	*Autographa pulchrina*	11968	100.00%	0	94.91
32	4593	11.2	5	circular	*Eupsilia transversa*	436	18.00%	7.00E-117	76.22
33	4518	1.22	0	circular	*Eilema depressum*	4121	85.00%	0	85.86
34	4496	13.8	5	circular	*Hecatera dysodea*	1122	50.00%	0	75.81
35	4461	5.03	0	circular	*Phalera bucephala*	5858	99.00%	0	90.37
36	4437	22.59	5	circular	*Lymantria monacha*	4368	92.00%	0	86.02
37	4334	4.85	5	circular	*Pyrgus malvae*	1788	83.00%	0	75.85
38	4314	3.78	5	circular	*Agrochola circellaris*	2704	90.00%	0	79.34
39	3869	2.03	0	circular	*Enterococcus faecium*	198	6.00%	3.00E-45	81.42
40	3494	10.78	0	circular	*Hesperia comma*	468	9.00%	2.00E-126	91.57
41	3268	2.77	5	circular	*Atethmia centrago*	783	22.00%	1.00E-63	86.42
42	2695	19.95	5	linear	*Helicoverpa armigera*	130	3.00%	8.00E-25	88.68
43	2372	1.14	0	circular	*Noctua pronuba*	722	19.00%	9.00E-54	91.12
44	2144	3.13	5	circular	*Heliothis subflexa*	382	17.00%	6.00E-45	81.16
45	2001	1.13	0	circular	*Leptidea sinapis*	1290	75.00%	0	82.32
46	1991	3.4	0	circular	*Heliothis subflexa*	707	22.00%	4.00E-131	94.82
47	1899	1.86	0	circular	*Timema poppe*	56.5	1.00%	0.01	94.44
48	1796	2.49	5	circular	No significant similarity	-	-	-	-
49	1596	1.65	5	circular	*Heliothis subflexa*	1703	38.00%	2.00E-127	95.56
50	1426	0.88	5	circular	No significant similarity	-	-	-	-
51	1414	1.65	0	circular	*Autographa pulchrina*	8063	98.00%	0	96.37
52	1300	1.29	5	circular	*Mamestra brassicae*	409	20.00%	4.00E-75	89.74
53	1136	67.21	5	linear	*Heliothis subflexa*	580	21.00%	2.00E-112	97.93

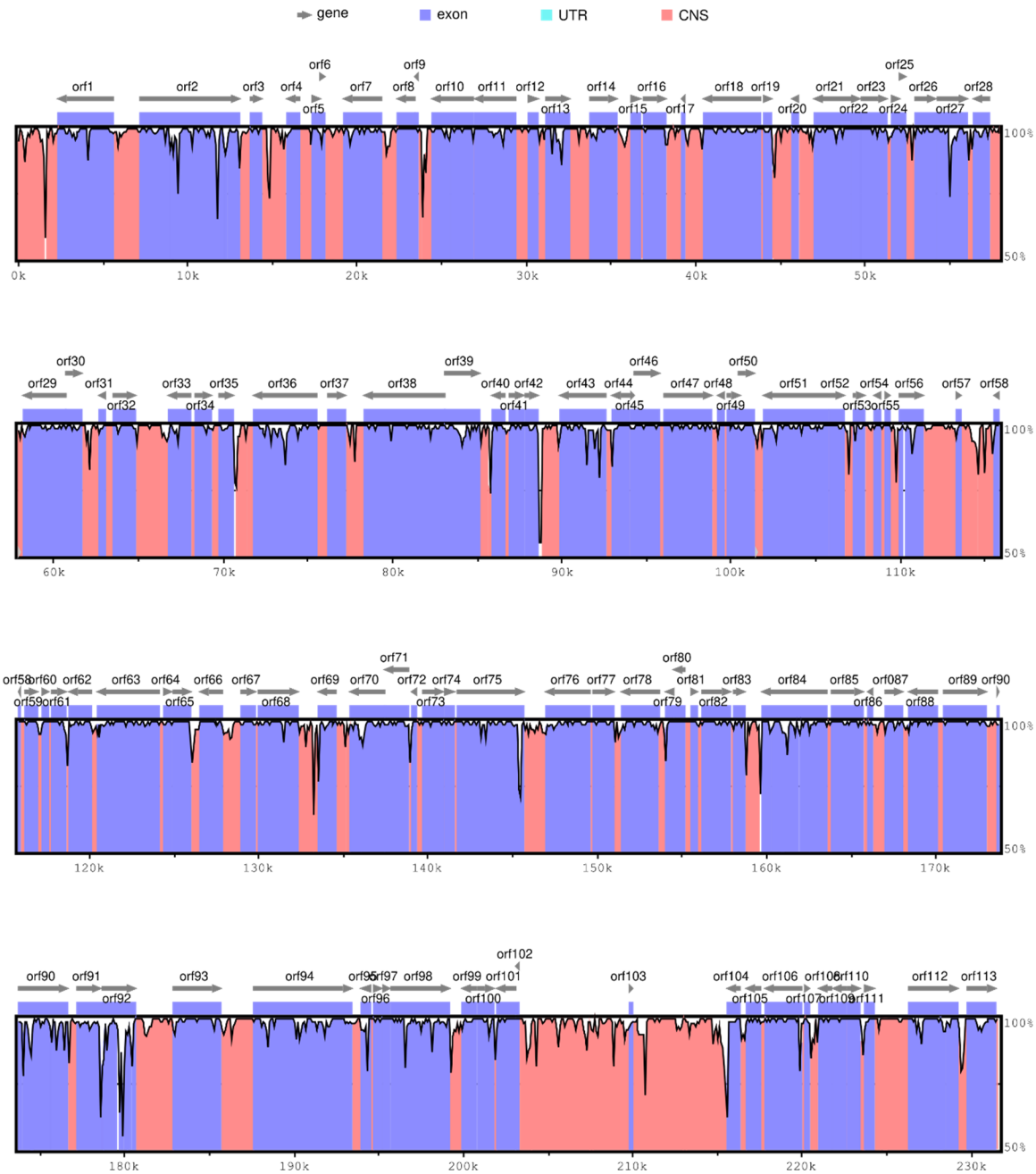
De novo analysis of our LA73 dataset with SPAdes software produced 53 different contigs. Three of these contigs showed sequence similarity to HzNV-2 genome and remaining contigs were aligned to distant lepidopteran species in BLAST search.

Figure 4.1: Whole genome alignment of HvNV and HzNV-2 genomes showing synteny blocks and sequence linearization



A) Pairwise alignment of HvNV (upper) and HzNV-2 (lower) genomes showing the locations in synteny blocks (green and red). These blocks indicate that circular HvNV and HzNV-2 genomes linearized from different locations. B) Once these circular genomes are linearized from the same location, they showed identical syntenies. No genomic rearrangements or sequence editing was performed on these sequnces beyond this adjustment.

Figure 4.2: Whole genome alignment of HvNV main contig and HzNV-2 genome. Bottom line represents the HzNV-2 genome and the top line represents the HvNV genome. Nucleotide variations are indicated as vertical lines and gaps. ORF distribution of HzNV-2 genome shown on top and the sequence simllarity is represented at 100 bp resolution ranging from 50% to 100%.



UTR: Untranslated region, CNS: Conserved non-coding region.

Table 4.3: List of all HzNV-2 open reading frames found in HvNV genome

Gene	Direction	Length in HzNV-2	Length	Length (w/ gaps)	Sequence Similarity	Product
orf1	reverse	3336	3345	3427	96.71%	histone H3
orf2	forward	5868	5754	6019	89.24%	hypothetical protein
orf3	forward	705	705	705	99.15%	hypothetical protein
orf4	reverse	768	768	768	98.96%	hypothetical protein
orf5	forward	441	441	441	98.19%	hypothetical protein
orf6	forward	288	219	235	91.67%	hypothetical protein
orf7	reverse	2166	2193	2198	96.99%	juvenile hormone esterase
orf8	reverse	1008	1008	1008	99.60%	integrase
orf9	reverse	195	195	195	98.97%	hypothetical protein
orf10	reverse	2418	2436	2438	97.66%	hypothetical protein
orf11	reverse	2418	2424	2427	98.19%	hypothetical protein
orf12	forward	564	555	564	98.23%	inhibitor of apoptosis protein
orf13	forward	1347	1293	1408	90.59%	hypothetical protein
orf14	forward	1581	1563	1586	97.03%	hypothetical protein
orf15	forward	549	549	549	100.00%	inhibitor of apoptosis protein
orf16	forward	1326	1326	1326	99.70%	hypothetical protein
orf17	reverse	192	189	192	97.92%	hypothetical protein
orf18	reverse	3411	3420	3420	98.51%	DNA polymerase
orf19	forward	414	414	414	99.03%	hypothetical protein
orf20	reverse	432	432	432	99.31%	hypothetical protein
orf21	reverse	2079	2091	2091	97.85%	hypothetical protein
orf22	forward	525	525	525	99.81%	hypothetical protein
orf23	forward	1539	1539	1539	98.90%	membrane transporter
orf24	forward	429	429	429	99.07%	hypothetical protein
orf25	forward	330	330	330	99.39%	11k virion structural protein
orf26	forward	1143	1143	1143	98.78%	per-os infectivity factor 2
orf27	forward	1803	1755	1839	94.79%	hypothetical protein
orf28	reverse	1002	1017	1024	94.69%	very late expression factor 1
orf29	reverse	2472	2472	2472	98.58%	DNA repair related ATPase
orf30	forward	969	969	969	98.66%	hypothetical protein
orf31	reverse	396	396	396	98.23%	hypothetical protein
orf32	forward	1275	1275	1279	99.22%	hypothetical protein
orf33	reverse	1287	1281	1287	97.82%	hypothetical protein
orf34	forward	1008	1014	1014	97.73%	guanosine monophosphate kinase
orf35	forward	876	876	876	98.40%	thymidylate synthase
orf36	reverse	3771	3744	3823	93.80%	hypothetical protein
orf37	forward	1023	1023	1023	99.80%	hypothetical protein
orf38	reverse	4746	4764	4784	98.14%	helicase
orf39	forward	2112	2082	2135	94.37%	baculovirus 19k protein
orf40	reverse	726	726	726	99.17%	late expression factor 5
orf41	forward	897	897	897	99.00%	hypothetical protein
orf42	forward	753	762	762	97.24%	28k virion structural protein
orf43	reverse	2622	2685	2814	94.50%	late expression factor 4
orf44	reverse	1044	1044	1044	99.04%	hypothetical protein
orf45	forward	207	219	219	91.78%	hypothetical protein

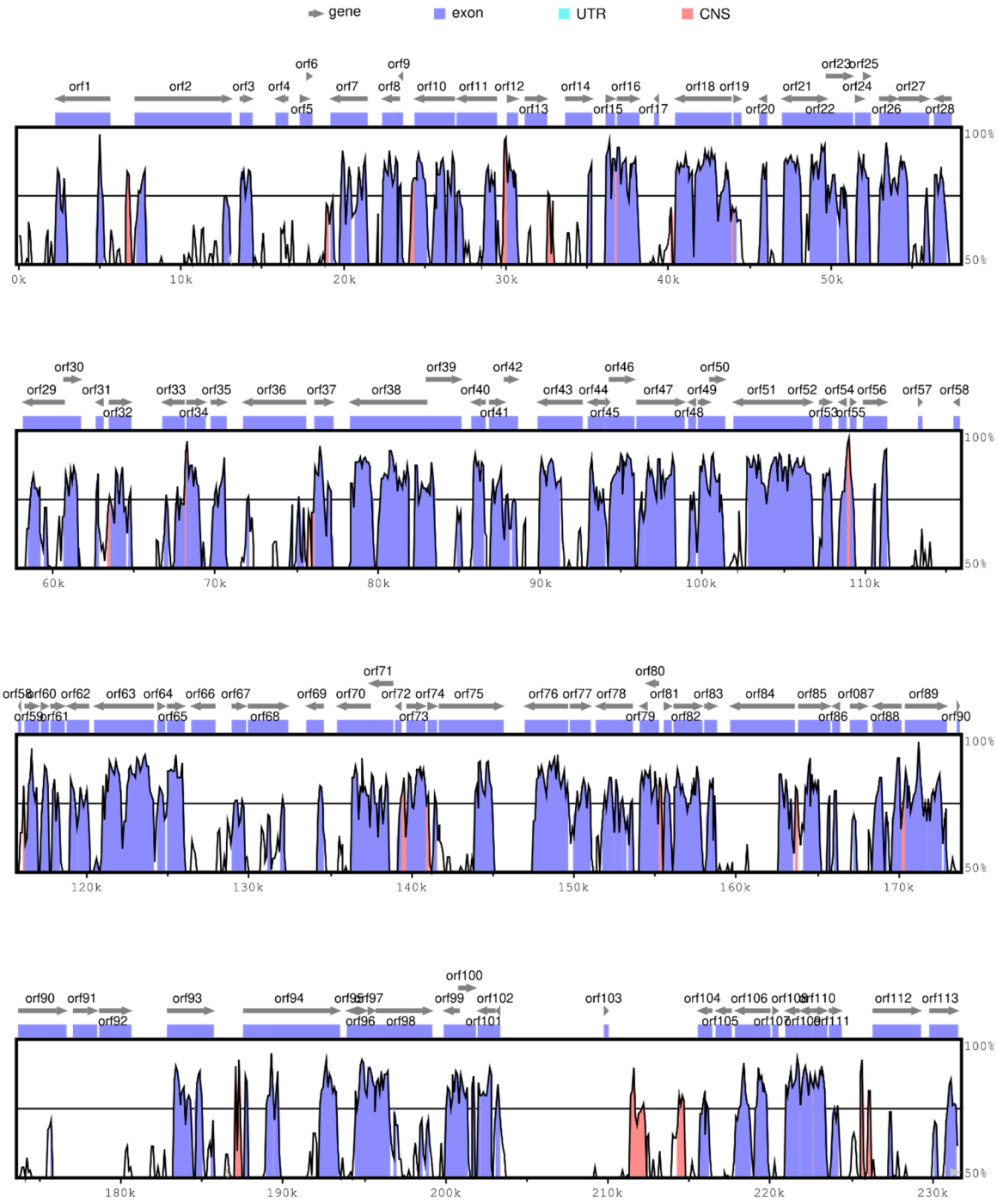
Table 4.3 (continued)

orf46	forward	1473	1473	1473	99.12%	hypothetical protein
orf47	forward	2817	2820	2820	98.90%	ribonuclease reductase
orf48	reverse	363	366	366	98.91%	hypothetical protein
orf49	forward	747	747	747	99.46%	hypothetical protein
orf50	forward	882	879	882	98.41%	hypothetical protein
orf51	reverse	3798	3789	3798	98.31%	late expression factor 8
orf52	forward	846	846	846	99.17%	31k virion structural protein
orf53	forward	645	648	648	97.99%	per-os infectivity factor 3
orf54	reverse	444	444	444	99.55%	hypothetical protein
orf55	forward	285	285	285	99.65%	hypothetical protein
orf56	forward	1416	1407	1433	95.46%	hypothetical protein
orf57	forward	216	177	177	98.62%	hypothetical protein
orf58	reverse	417	294	294	98.33%	hypothetical protein
orf59	forward	765	765	765	99.08%	p34 late protein
orf60	forward	375	375	375	99.47%	hypothetical protein
orf61	forward	771	750	775	96.24%	hypothetical protein
orf62	reverse	1323	1323	1323	98.49%	odv-e56 structural protein
orf63	reverse	3612	3624	3653	98.35%	late expression factor 9
orf64	forward	465	465	465	99.14%	hypothetical protein
orf65	forward	1002	1002	1002	99.50%	ribonuclease reductase
orf66	reverse	1326	1326	1326	99.40%	serine hydroxymethyltransferase
orf67	forward	825	825	825	99.03%	deoxynuclease kinase
orf68	forward	2370	2370	2379	98.07%	matrix metalloprotease
orf69	reverse	1053	1068	1075	98.01%	dUTPase
orf70	reverse	1974	1959	1986	95.72%	DNA excision repair enzyme
orf71	reverse	1389	1389	1389	98.78%	hypothetical protein
orf72	reverse	339	339	339	98.82%	hypothetical protein
orf73	forward	1152	1152	1152	99.22%	p51 late protein
orf74	forward	555	555	555	99.28%	hypothetical protein
orf75	forward	3948	3915	4029	91.80%	hypothetical protein
orf76	reverse	2631	2631	2631	98.71%	hypothetical protein
orf77	forward	1233	1233	1233	99.11%	hypothetical protein
orf78	reverse	2190	2178	2190	97.81%	hypothetical protein
orf79	reverse	438	438	438	99.77%	hypothetical protein
orf80	reverse	714	714	714	99.44%	hypothetical protein
orf81	forward	384	384	384	99.48%	hypothetical protein
orf82	forward	1707	1707	1707	99.18%	per-os infectivity factor 1
orf83	forward	666	666	666	99.25%	hypothetical protein
orf84	reverse	3867	3900	4047	74.61%	hypothetical protein
orf85	forward	1902	1902	1902	99.11%	hypothetical protein
orf86	reverse	348	363	363	97.99%	hypothetical protein
orf87	forward	969	972	972	98.35%	hypothetical protein
orf88	reverse	1659	1665	1665	98.68%	hypothetical protein
orf89	forward	2496	2496	2496	99.48%	vp91 capsid protein

Table 4.3 (continued)

orf90	forward	2925	2961	3100	79.37%	hypothetical protein
orf91	forward	1392	1369	1409	89.73%	hypothetical protein
orf92	forward	1941	1980	2223	77.99%	hypothetical protein
orf93	forward	2847	2841	2862	97.69%	methyltransferase
orf94	forward	5862	5856	5960	84.94%	DNA ligase
orf95	reverse	588	675	686	83.92%	hypothetical protein
orf96	forward	471	471	471	99.15%	Ac81-like protein
orf97	forward	402	402	402	99.25%	15k virion structural protein
orf98	forward	3438	3372	3443	95.90%	hypothetical protein
orf99	reverse	867	867	867	99.19%	esterase
orf100	forward	942	939	942	97.66%	hypothetical protein
orf101	reverse	1071	1071	1071	99.35%	hypothetical protein
orf102	reverse	210	210	210	100.00%	hypothetical protein
orf103	forward	201	228	231	97.01%	hypothetical protein
orf104	reverse	735	735	735	99.05%	hypothetical protein
orf105	reverse	879	879	879	98.41%	hypothetical protein
orf106	reverse	2094	2055	2097	97.09%	p74 envelope protein
orf107	forward	291	276	278	89.13%	hypothetical protein
orf108	reverse	780	780	780	99.36%	baculovirus 38k protein
orf109	reverse	732	732	732	99.32%	hypothetical protein
orf110	forward	765	765	765	99.35%	protein kinase
orf111	forward	645	645	645	99.07%	dihydrofolate reductase
orf112	forward	2895	3063	3198	94.31%	hypothetical protein
orf113	forward	1689	1734	1734	94.98%	hypothetical protein

Figure 4.3: Whole genome alignment of the defective *H. virescens* nudivirus contig against HzNV-2 genome. Bottom line represents the HzNV-2 genome and the top line represents the defective HvNV genome. ORF distribution of HzNV-2 genome shown on top and the sequence simllarity is represented at 100 bp resolution ranging from 50% to 100%.



UTR: Untranslated region, CNS: Conserved non-coding region.

Table 4.4: List of all HzNV-2 open reading frames found in defective nudivirus genome. Genes in red at end of table have been deleted in the HvNV defective genome.

Gene	Direction	Length in HzNV-2	Length	Length (w/ gaps)	Sequence Similarity	Product
orf3	forward	705	705	705	73.76%	hypothetical protein
orf7	reverse	2166	1941	3205	58.26%	juvenile hormone esterase
orf8	reverse	1008	924	1075	73.61%	integrase
orf9	reverse	195	195	711	69.23%	hypothetical protein
orf10	reverse	2418	2118	2673	51.23%	hypothetical protein
orf12	forward	564	534	534	78.67%	inhibitor of apoptosis protein
orf15	forward	549	534	549	61.57%	inhibitor of apoptosis protein
orf16	forward	1326	1299	3114	76.70%	hypothetical protein
orf18	reverse	3411	3312	3544	77.48%	DNA polymerase
orf19	forward	414	495	1726	48.99%	hypothetical protein
orf20	reverse	432	432	435	77.24%	hypothetical protein
orf21	reverse	2079	2068	2625	58.91%	hypothetical protein
orf22	forward	525	525	585	87.24%	hypothetical protein
orf23	forward	1539	1392	1855	62.98%	membrane transporter
orf24	forward	429	429	429	82.52%	hypothetical protein
orf25	forward	330	330	330	77.88%	11k virion structural protein
orf26	forward	1143	1143	1144	74.98%	per-os infectivity factor 2
orf27	forward	1803	1788	3178	46.03%	hypothetical protein
orf28	reverse	1002	852	2028	66.57%	very late expression factor 1
orf29	reverse	2472	2466	3150	46.38%	DNA repair related ATPase
orf30	forward	969	795	1999	66.91%	hypothetical protein
orf31	reverse	396	336	405	53.77%	hypothetical protein
orf32	forward	1275	1224	3325	64.72%	hypothetical protein
orf33	reverse	1287	1104	1424	48.19%	hypothetical protein
orf34	forward	1008	924	1503	67.92%	guanosine monophosphate kinase
orf35	forward	876	864	2093	73.97%	thymidylate synthase
orf37	forward	1023	930	1114	69.70%	hypothetical protein
orf38	reverse	4746	4726	5562	47.71%	helicase
orf40	reverse	726	837	1242	78.51%	late expression factor 5
orf41	forward	897	885	915	63.36%	hypothetical protein
orf42	forward	753	714	1151	58.96%	28k virion structural protein
orf43	reverse	2622	2636	4285	54.53%	late expression factor 4
orf44	reverse	1044	1023	1026	70.48%	hypothetical protein
orf46	forward	1473	1458	1733	81.40%	hypothetical protein
orf47	forward	2817	2835	3694	64.94%	ribonuclease reductase

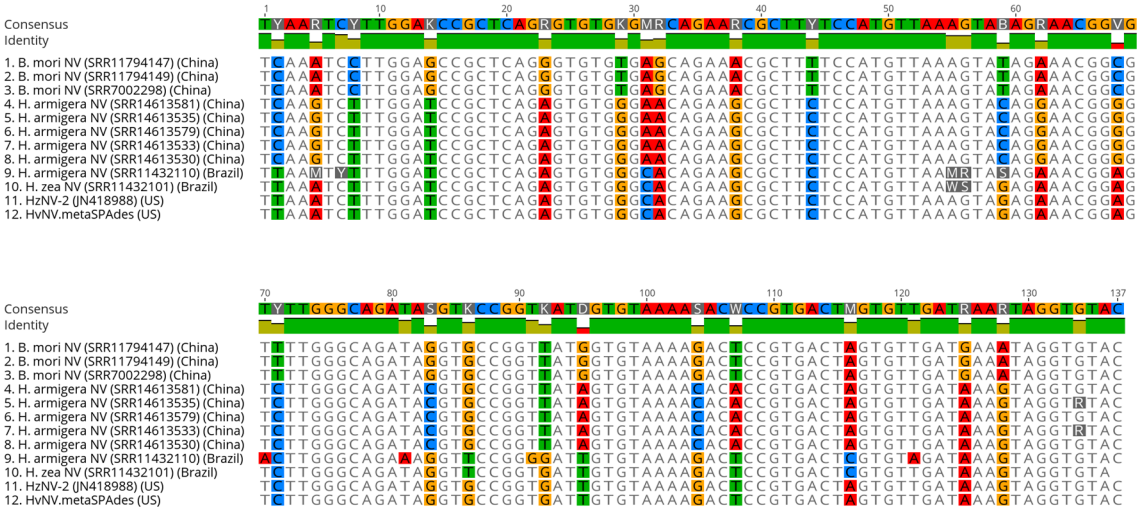
Table 4.4 (continued)

orf48	reverse	363	375	379	68.75%	hypothetical protein
orf49	forward	747	735	735	79.52%	hypothetical protein
orf50	forward	882	849	1873	71.19%	hypothetical protein
orf51	reverse	3798	3774	5143	70.52%	late expression factor 8
orf52	forward	846	846	846	83.92%	31k virion structural protein
orf53	forward	645	618	1139	73.33%	per-os infectivity factor 3
orf54	reverse	444	441	441	71.40%	hypothetical protein
orf55	forward	285	261	1455	60.70%	hypothetical protein
orf59	forward	765	759	1033	75.65%	p34 late protein
orf60	forward	375	378	680	79.37%	hypothetical protein
orf61	forward	771	639	1066	51.36%	hypothetical protein
orf62	reverse	1323	1263	1271	65.99%	odv-e56 structural protein
orf63	reverse	3612	3531	3740	72.15%	late expression factor 9
orf64	forward	465	462	467	69.89%	hypothetical protein
orf65	forward	1002	1002	1552	82.04%	ribonuclease reductase
orf67	forward	825	771	1836	59.08%	deoxynuclease kinase
orf69	reverse	1053	1083	1359	50.05%	dUTPase
orf70	reverse	1974	1979	3586	55.10%	DNA excision repair enzyme
orf71	reverse	1389	1194	1839	59.99%	hypothetical protein
orf73	forward	1152	1152	1198	78.04%	p51 late protein
orf75	forward	3948	3920	7573	45.12%	hypothetical protein
orf76	reverse	2631	2613	3381	71.34%	hypothetical protein
orf77	forward	1233	1197	1694	68.90%	hypothetical protein
orf78	reverse	2190	1998	2766	60.04%	hypothetical protein
orf79	reverse	438	432	460	69.47%	hypothetical protein
orf80	reverse	714	699	721	73.39%	hypothetical protein
orf81	forward	384	378	598	74.74%	hypothetical protein
orf82	forward	1707	1698	2027	74.14%	per-os infectivity factor 1
orf83	forward	666	663	1667	62.03%	hypothetical protein
orf85	forward	1902	1887	4980	59.41%	hypothetical protein
orf86	reverse	348	356	2627	53.65%	hypothetical protein
orf88	reverse	1659	1521	1945	46.16%	hypothetical protein
orf89	forward	2496	2511	3968	66.43%	vp91 capsid protein
orf92	forward	1941	1969	1969	45.63%	hypothetical protein
orf93	forward	2847	2822	4825	46.87%	methyltransferase
orf95	reverse	588	738	817	45.07%	hypothetical protein
orf96	forward	471	471	471	81.95%	Ac81-like protein

Table 4.4 (continued)

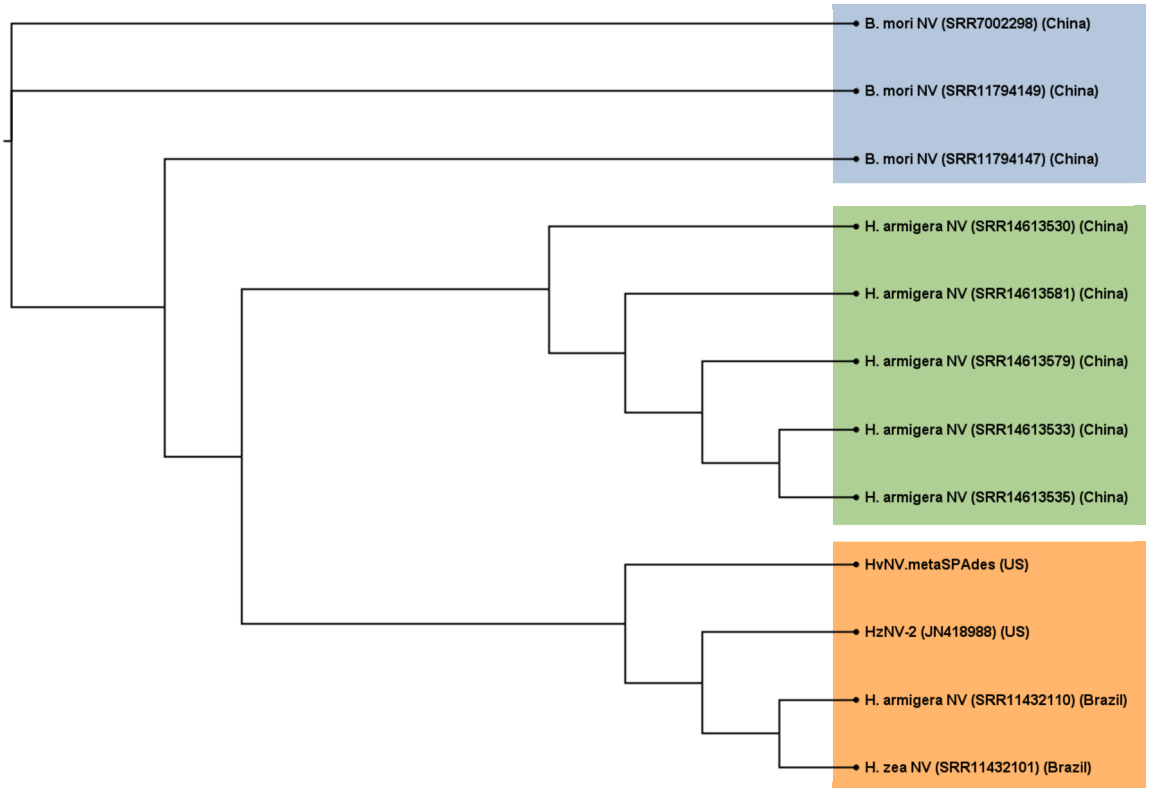
orf97	forward	402	414	423	78.49%	15k virion structural protein
orf99	reverse	867	852	867	74.05%	esterase
orf100	forward	942	759	1228	54.35%	hypothetical protein
orf101	reverse	1071	1086	1300	62.65%	hypothetical protein
orf102	reverse	210	219	219	73.06%	hypothetical protein
orf104	reverse	735	738	2308	66.53%	hypothetical protein
orf106	reverse	2094	2127	2293	51.19%	p74 envelope protein
orf108	reverse	780	780	780	83.08%	baculovirus 38k protein
orf109	reverse	732	735	739	81.36%	hypothetical protein
orf110	forward	765	768	1094	82.75%	protein kinase
orf111	forward	645	564	1321	58.90%	dihydrofolate reductase
orf113	forward	1762	1748	2382	46.86%	hypothetical protein
orf1	reverse	-	-	-	-	histone H3
orf4	reverse	-	-	-	-	hypothetical protein
orf6	reverse	-	-	-	-	hypothetical protein
orf13	forward	-	-	-	-	hypothetical protein
orf14	forward	-	-	-	-	hypothetical protein
orf17	reverse	-	-	-	-	hypothetical protein
orf36	reverse	-	-	-	-	hypothetical protein
orf39	forward	-	-	-	-	baculovirus 19k protein
orf45	forward	-	-	-	-	hypothetical protein
orf56	forward	-	-	-	-	hypothetical protein
orf57	reverse	-	-	-	-	hypothetical protein
orf58	forward	-	-	-	-	hypothetical protein
orf66	forward	-	-	-	-	serine hydroxymethyltransferase
orf68	reverse	-	-	-	-	matrix metalloprotease
orf72	forward	-	-	-	-	hypothetical protein
orf74	reverse	-	-	-	-	hypothetical protein
orf84	forward	-	-	-	-	hypothetical protein
orf87	forward	-	-	-	-	hypothetical protein
orf91	forward	-	-	-	-	hypothetical protein
orf94	forward	-	-	-	-	DNA ligase
orf98	forward	-	-	-	-	hypothetical protein
orf103	forward	-	-	-	-	hypothetical protein
orf105	reverse	-	-	-	-	hypothetical protein
orf107	forward	-	-	-	-	hypothetical protein
orf112	forward	-	-	-	-	hypothetical protein

Figure 4.4: Multiple sequence alignment of fragments homologous to HzNV2-ORF38 found in heliothine nudiviruses.



The ORF38 of HzNV-2 genome had enough sequence data and spatial diversity to perform phylogenetic analysis. A multiple sequence alignment block was generated using 10 homologus sequences from those deposited by researchers from China, Brazil and the US. Color-coded boxes represent nucleotide polymorphisms (Red: Adenosine, Green: Thymine, Blue: Cytosine, Orange: Guanine)

Figure 4.5: Phylogenetic tree generated based on the multiple sequence alignment shown in Fig. 4.3.



Phylogenetic relations reconstructed based on available sequence data which is homologous to HzNV2-ORF38 gene. Viral sequences shown in both green and blue boxes cluster in parallel with host species, however, samples in orange box shows multi-species grouping.

CHAPTER 5: EPILOGUE AND DISCUSSION

5.1 Epilogue

This project evolved remarkably since its first proposal. Our initial research questions have changed several times throughout my doctoral work due to experimental failures, unexpected circumstances and partly because of the COVID-19 pandemic. Besides many other small scale experiments related to nudivirus pathology, a major failure in HvNV propagation experiments led me to use public databases and bioinformatic tools for investigating other aspects of Heliothine nudiviruses. Unlike Baculoviruses, Heliothine nudiviruses do not form protective occlusion bodies and are fairly susceptible to environmental factors even in insect tissues. From a technical stand point, HvNV propagation was achievable despite the fact that the samples we processed were not collected specifically for our study. First, we tested our paper-smear samples for propagation experiments by preparing inoculates from known nudivirus positive samples that were collected from *H. virescens* host. At that point, we didn't have any procedures to verify the host species so we used the information provided by the field researcher. After several attempts with HvNV inoculates, we were not able to propagate the virus neither in Sf-9 cell lines nor in its native host, *Heliothis virescens* insects. This failure was primarily caused by the methods used in sample collection, storage and shipping that were not suitable for live nudivirus isolation. In the following year we obtained more than 200 frozen *H. virescens* moths collected from a research station in Louisiana with pheromone traps so there were no female moths in this batch. These samples were also exposed to excessive heat and UV light, however we assumed that the virus can be recovered from insect tissues since it provides a cellular protective layer for some time. This assumption was made based

on our observations with *H. zea* nudivirus, which can survive even if the infected insect body stored in room temperatures for a few days. So, we treated all samples as nudivirus positive and prepared inoculates from their abdomens using sterilized utensils and sterile PBS solution, which is commonly used in HzNV experiments. Small portions of these inoculates were used for DNA extraction and PCR screening, however only 11 samples showed nudivirus positives. Two of these positives (LA27 and LA73) exhibited very strong banding pattern in gel electrophoresis so individual inoculates were prepared from LA27 and LA73. Along with other nudivirus positives which were pooled in pairs, we prepared a total of 6 injection-ready solutions by filter-sterilization.

We used these sterile inoculates to infect Sf-9 cells in two replicates however no cytopathic effects were observed in serial passages. Next, we injected inoculates into 24 newly emerged *H. virescens* abdomens in six replicates. No morphological changes were observed in these infected insects after 5 days and only a small percentage of them showed nudivirus presence in PCR assays. Eggs from infected females were reared for two generations; only a few nudivirus positives found in the first generation and none in the second generation. These results indicated that the nudivirus titers were either too low or there were no viable virus particles to establish an infection. In conjunction with sampling method, we concluded HvNV virions are not capable of surviving the conditions in a regular trap capture setting.

Finally, in the following year, we received 79 live capture *H. virescens* samples from another research station in Louisiana. In this batch, insects were frozen directly after capture and stored in -20°C freezer until it was shipped in to our laboratory in dry ice. The number of samples dropped significantly due to COVID-19 regulations and only three

nudivirus positives were detected in this second batch. Despite the fact that the positives in this batch were fairly weak, we prepared inoculates again and injected them in 24 *H. virescens* abdomens in 2 replicates. Similar to previous experiments, no nudivirus replication was detected in virus injected insects.

In conclusion, we suggest two strategies to researchers that are planning to isolate and propagate Heliothine nudiviruses. In the scenario, the probability of isolating live nudivirus can be dramatically improved by increasing the number of live-capture insects. These insects must be transferred to a research lab in cold chain to prevent any degradation in live virus titers. We think that such a study should aim to collect at least several hundred insects to acquire viable virions. Another strategy to achieve a successful nudivirus propagation would be rearing insects collected from source population for several generations until viral pathology is detected in morphology or by molecular screening methods. Undoubtedly, this strategy requires much more time and resources, which further increases the overall cost of the study.

5.2 Discussion

Bollworms are a group of lepidopteran insects that feed on fruiting body at larval stage. The bollworm complex contains major agricultural pests that cause severe damage many economically important crops and horticulture products. Three bollworm species, *Helicoverpa zea, Heliothis virescens, and Helicoverpa armigera* are among the costliest crop pests in the world (Hardwick, 1965). These insect are highly polypagous and in the larval stage, they enter into fruiting bodies of the plant and feed on vegetative and reproductive tissues in plants such as corn, cotton, lettuce, soybean, tomato and tobacco,

and many others. Many conventional insecticides exhibit poor performance in late instars due to the protected micro-environment provided by the fruiting body.

Besides their role in spreading many vector-borne diseases, insects are also susceptible to many bacterial and viral pathogens. Baculoviruses are well studied viruses and in many cases they serve as an effective biopesticide. They are also engineered to to produce recombinant proteins at larger scales in insect cell based bioreactors. Nudiviridae family is ancestral to Baculoviridae and Bracoviridae families (Thézé et al., 2011) and it contains non-occluded viruses that can infect several insect and arthropod species. Helicoverpa zea nudivirus – 2 (HzNV-2) is a sexually transmitted nudivirus that cause gonad trophy in *H. zea* (corn earworm) adults. It also exhibits a biphasic replication pattern where the infection is asymptomatic in latent phase and productive in lytic phase. The productive phase is characterized by gonad atrophy and a visible viral plug formation (Burand & Lu, 1997; Rallis & Burand, 2002). Almost 1/3 of all HzNV-2 infections are lytic that cause agonadal sterility due to physiological and morphological defects in host reproductive tracts (Raina & Adams, 1995). Another symptom of the infection is the elevated pheromone synthesis and continual mating calls in females (Burand et al., 2005; Burand & Tan, 2006). Along with viral plug formation, changes in mating behavior help the virus to spread horizontally relatively quickly. In addition to horizontal transmission, HzNV-2 can also be transmitted vertically through generations. A cell line restricted strain of this virus, HzNV-1 was discovered in 1971 and its genome was sequenced in 2005 which makes it the first known and sequenced Heliothine nudivirus. Genome of HzNV-1 closely resembles the HzNV-2 in terms of sequence identity, however, the former has several

indels and missing genes. Based similarities between two nudivirus, HzNV-1 has potentially derived from HzNV-2 during cell immortalization process.

In lytic phase, HzNV-2 infection cause host tissue atrophy in both female and male gonads and thus sterilize the host. This agonadal phenotype is found only in 1/3 of all infected population. Also, HzNV-2 modifies the host behavior in females by inhibiting the activity of a pheromonostatic peptide which is responsible for ending mating calls. All these pathological features make the HzNV-2 a candidate for corn earworm management however, early studies with HzNV-1 strain pointed that persistently infected TN-368 cell lines can be resistant to superinfection (Ralston et al., 1981). Even though HzNV-2 is a relatively well-studied virus, its prevalence in wild corn earworm populations is largely unknown. Here in this project, we have investigated HzNV-2 distribution in the Cotton Belt region and several other surrounding states. Additionally, we performed a digital survey on available genetic databases in using bioinformatic tools. Finally, we sequenced a novel nudivirus that infects *H. virescens* populations along with a defective viral genome.

Results of this study indicate that nudiviruses are circulating in at least three bollworm species around the world; *H. zea, H. armigera,* and *H. virescens*. Corn earworm (*H. zea*) lineage diverged from the Old World bollworm (*H. armigera*) populations and established in Americas 1.5 Mya (Pearce et al., 2017b). As a result of this divergence, heterozygosity in *H. zea* populations are significantly lower than both *H. armigera* and *H. punctigera* populations (Mallet et al., 1993; Seymour et al., 2016). Our data suggest that HzNV-2 prevalence in the U.S. originates from source populations in South America where *H. zea* and *H. armigera* populations coexist and hybridize in many different habitats (Cordeiro et al., 2020). This hypothesis is also supported by SNP percentages which are

considerably higher in *H. armigera* nudivirus compared to *H. zea* nudivirus found in Brazil populations (Fig. 3.4). Furthermore, it can be argued that HzNV is an evolutionary lineage diverged from an ancestral *H. armigera* nudivirus and shows signs of a genetic bottleneck similar to its native host. In this study, we also present a series of strong evidences that indicates nudivirus prevalence in Greece, China and Australia. This evidence indicates that Heliothine nudiviruses in Americas may originally diverged from an ancestor found in the Old World. This claim is also supported by the sequences found in many experiments based on Bombyx mori cell lines conducted in China. Finally, we found traces of nudivirus DNA in several other species including S*podoptera frugiperda*, *Helicoverpa assulta*, and *Ostrinia nubilalis*, however, the number sequences were not enough to make any inferences. Traces of nudiviral sequences may indicate a very low level infection or endogenous viral sequences which also supports the idea of presence of specific nudiviruses for these host species.

In addition to the diversity of hosts that exhibit nudivirus presence, we also found a significant difference in terms of host species and read numbers when we process datasets using either HzNV-1 and HzNV-2 reference genomes. When the datasets aligned to HzNV-1 genome, we found nudivirus less amount of short reads from a wider range of host species. Conversely, when we align the same datasets against HzNV-2 reference genome, we found more short reads from less number host species. Since HzNV-1 is a cell line restricted nudivirus, it is not found in feral corn earworm populations. On the other hand, the difference between known HzNV-1 and HzNV-2 genomes either indicates a large HzNV quasispecies circulating in feral Heliothine populations or supports the idea of multiple historic transposon activity.

Finally, we isolated and sequenced a novel nudivirus from an infected tobacco budworm (*H. virescens*) sample collected from Louisiana. This new nudivirus, HvNV, closely resembles HzNV-2 virus in terms of genomic structure and sequence similarity. Since *H. zea* and *H. virescens* popluations overlap in Cotton Belt region, it is possible that HzNV was transmitted to *H. virescens* host as a result of interspecies mating attempts between infected *H. zea* and healthy *H. virescens*. We also found a smaller homologous nudivirus genome with multiple large deletions. Defective interfereing particles (DIPs) are generated in many viral infections including Baculoviruses, the particle I found resembles a defective viral genome (DVG) and it is possibly playing a role in HvNV pathology.

In conclusion, our results support the idea of a large and diverse group of lepidopteran nudiviruses circulating in natural populations, potentially transmitting sexually and exhibiting various degrees of reproductive sterility. If this is true, then it may be possible to develop a series of nudiviral insecticides for these costly and resistant pest species.

REFRENCES

Achaleke, J., Martin, T., Ghogomu, R. T., Vaissayre, M., & Brévault, T. (2009). Esterase-mediated resistance to pyrethroids in field populations of Helicoverpa armigera (Lepidoptera: Noctuidae) from Central Africa: Resistance to pyrethroids in H. armigera in Central Africa. Pest Management Science, 65(10), 1147–1154. https://doi.org/10.1002/ps.1807

Achee, N. L., Grieco, J. P., Vatandoost, H., Seixas, G., Pinto, J., Ching-NG, L., Martins, A. J., Juntarajumnong, W., Corbel, V., Gouagna, C., David, J.-P., Logan, J. G., Orsborne, J., Marois, E., Devine, G. J., & Vontas, J. (2019). Alternative strategies for mosquito-borne arbovirus control. PLOS Neglected Tropical Diseases, 13(1), e0006822. https://doi.org/10.1371/journal.pntd.0006822

Adams, B. P., Catchot, A. L., Cook, D. R., Gore, J., Musser, F. R., Irby, J. T., & Golden, B. R. (2015). The Impact of Simulated Corn Earworm (Lepidoptera: Noctuidae) Damage in Indeterminate Soybean. Journal of Economic Entomology, 108(3), 1072–1078. https://doi.org/10.1093/jee/tov094

Adams, B. P., Cook, D. R., Catchot, A. L., Gore, J., Musser, F., Stewart, S. D., Kerns, D. L., Lorenz, G. M., Irby, J. T., & Golden, B. (2016). Evaluation of Corn Earworm, *Helicoverpa zea* (Lepidoptera: Noctuidae), Economic Injury Levels in Mid-South Reproductive Stage Soybean. Journal of Economic Entomology, 109(3), 1161–1166. https://doi.org/10.1093/jee/tow052

Alnaji, F. G., Holmes, J. R., Rendon, G., Vera, J. C., Fields, C. J., Martin, B. E., & Brooke, C. B. (2019). Sequencing Framework for the Sensitive Detection and Precise Mapping of Defective Interfering Particle-Associated Deletions across Influenza A and B Viruses. Journal of Virology, 93(11), e00354-19. https://doi.org/10.1128/JVI.00354-19

Anderson, C. J., Oakeshott, J. G., Tay, W. T., Gordon, K. H. J., Zwick, A., & Walsh, T. K. (2018). Hybridization and gene flow in the mega-pest lineage of moth, *Helicoverpa*. Proceedings of the National Academy of Sciences, 115(19), 5034–5039. https://doi.org/10.1073/pnas.1718831115

Antipov, D., Raiko, M., Lapidus, A., & Pevzner, P. A. (2020). MetaviralSPAdes: Assembly of viruses from metagenomic data. Bioinformatics, 36(14), 4126–4129. https://doi.org/10.1093/bioinformatics/btaa490

Bailey, L., Carpenter, J. M., & Woods, R. D. (1981). Properties of a filamentous virus of the honey bee (*Apis mellifera*). Virology, 114(1), 1–7. https://doi.org/10.1016/0042-6822(81)90247-6

Bateman, K. S., & Stentiford, G. D. (2017). A taxonomic review of viruses infecting crustaceans with an emphasis on wild hosts. Journal of Invertebrate Pathology, 147, 86–110. https://doi.org/10.1016/j.jip.2017.01.010

Beard, C. B., Butler, J. F., & Maruniak, J. E. (1989). A baculovirus in the flea, *Pulex simulans*. Journal of Invertebrate Pathology, 54(1), 128–131. https://doi.org/10.1016/0022-2011(89)90150-X

Beas-Catena, A., Sánchez-Mirón, A., García-Camacho, F., Contreras-Gómez, A., & Molina-Grima, E. (2014). Baculovirus biopesticides: An overview. JAPS, Journal of Animal and Plant Sciences, 24(2), 362–373.

Bedford, G. O. (1980). Biology, Ecology, and Control of Palm Rhinoceros Beetles. Annual Review of Entomology, 25(1), 309–339. https://doi.org/10.1146/annurev.en.25.010180.001521

Bedford, G. O. (2013). Long-term reduction in damage by rhinoceros beetle *Oryctes rhinoceros* (L.) (Coleoptera: Scarabaeidae: Dynastinae) to coconut palms at Oryctes Nudivirus release sites on Viti Levu, Fiji. African Journal of Agricultural Research, 8(49), 6422–6425. https://doi.org/10.5897/AJAR2013.7013

Bereczky, S., Mårtensson, A., Gil, J. P., & Färnert, A. (2005). Short report: Rapid DNA extraction from archive blood spots on filter paper for genotyping of *Plasmodium falciparum*. The American Journal of Tropical Medicine and Hygiene, 72(3), 249–251.

Bézier, A., Harichaux, G., Musset, K., Labas, V., & Herniou, E. A. (2017). Qualitative proteomic analysis of *Tipula oleracea* nudivirus occlusion bodies. Journal of General Virology, 98(2), 284–295. https://doi.org/10.1099/jgv.0.000661

Bibb, J. L., Cook, D., Catchot, A., Musser, F., Stewart, S. D., Leonard, B. R., Buntin, G. D., Kerns, D., Allen, T. W., & Gore, J. (2018). Impact of Corn Earworm (Lepidoptera: Noctuidae) on Field Corn (Poales: Poaceae) Yield and Grain Quality. Journal of Economic Entomology, 111(3), 1249–1255. https://doi.org/10.1093/jee/toy082

Bober, R., Azrielli, A., & Rafaeli, A. (2010). Developmental regulation of the pheromone biosynthesis activating neuropeptide-receptor (PBAN-R): Re-evaluating the role of juvenile hormone. Insect Molecular Biology, 19(1), 77–86. https://doi.org/10.1111/j.1365-2583.2009.00937.x

Bolling, B. G., Weaver, S. C., Tesh, R. B., & Vasilakis, N. (2015). Insect-Specific Virus Discovery: Significance for the Arbovirus Community. Viruses, 7(9), 4911–4928. https://doi.org/10.3390/v7092851

Borman, A. M., Linton, C. J., Miles, S.-J., Campbell, C. K., & Johnson, E. M. (2006). Ultra-rapid preparation of total genomic DNA from isolates of yeast and mould using Whatman FTA filter paper technology – a reusable DNA archiving system. Medical Mycology, 44(5), 389–398. https://doi.org/10.1080/13693780600564613

Boucias, D. G., Maruniak, J. E., & Pendland, J. C. (1989). Characterization of a non-occluded baculovirus (subgroup C) from the field cricket,Gryllus rubens. Archives of Virology, 106(1), 93–102. https://doi.org/10.1007/BF01311041

Brévault, T., Tabashnik, B. E., & Carrière, Y. (2015). A seed mixture increases dominance of resistance to Bt cotton in *Helicoverpa zea*. Scientific Reports, 5(1), 9807. https://doi.org/10.1038/srep09807

Burand, J. P., Kim, W., Afonso, C. L., Tulman, E. R., Kutish, G. F., Lu, Z., & Rock, D. L. (2012). Analysis of the Genome of the Sexually Transmitted Insect Virus Helicoverpa zea Nudivirus 2. Viruses, 4(12), 28–61. https://doi.org/10.3390/v4010028

Burand, J. P., & Lu, H. (1997). Replication of a Gonad-Specific Insect Virus in TN-368 Cells in Culture. Journal of Invertebrate Pathology, 70(2), 88–95. https://doi.org/10.1006/jipa.1997.4676

Burand, J. P., & Rallis, C. P. (2004). In vivo dose-response of insects to Hz-2V infection. Virology Journal, 1, 15. https://doi.org/10.1186/1743-422X-1-15

Burand, J. P., Stiles, B., & Wood, H. A. (1983). Structural and Intracellular Proteins of the Nonoccluded Baculovirus HZ-1. Journal of Virology, 46(1), 137–142.

Burand, J. P., & Tan, W. (2006). Mate Preference and Mating Behavior of Male *Helicoverpa zea* (Lepidoptera: Noctuidae) Infected with the Sexually Transmitted Insect Virus Hz-2V. Annals of the Entomological Society of America, 99(5), 969–973. https://doi.org/10.1603/0013-8746(2006)99[969:MPAMBO]2.0.CO;2

Burand, J. P., Tan, W., Kim, W., Nojima, S., & Roelofs, W. (2005). Infection with the insect virus Hz-2v alters mating behavior and pheromone production in female Helicoverpa zea moths. Journal of Insect Science, 5. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1283887/

Cabodevilla, O., Ibañez, I., Simón, O., Murillo, R., Caballero, P., & Williams, T. (2011). Occlusion body pathogenicity, virulence and productivity traits vary with transmission strategy in a nucleopolyhedrovirus. Biological Control, 56(2), 184–192. https://doi.org/10.1016/j.biocontrol.2010.10.007

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. BMC Bioinformatics, 10, 421. https://doi.org/10.1186/1471-2105-10-421

Caron, A. W., Archambault, J., & Massie, B. (1990). High-level recombinant protein production in bioreactors using the baculovirus–insect cell expression system. Biotechnology and Bioengineering, 36(11), 1133–1140. https://doi.org/10.1002/bit.260361108

Carrière, Y., Degain, B., Unnithan, G. C., Harpold, V. S., Li, X., & Tabashnik, B. E. (2019). Seasonal Declines in Cry1Ac and Cry2Ab Concentration in Maturing Cotton Favor Faster Evolution of Resistance to Pyramided Bt Cotton in *Helicoverpa zea* (Lepidoptera: Noctuidae). Journal of Economic Entomology, 112(6), 2907–2914. https://doi.org/10.1093/jee/toz236

Chao, Y. C., Wood, H. A., Chang, C. Y., Lee, H. J., Shen, W. C., & Lee, H. T. (1992). Differential expression of Hz-1 baculovirus genes during productive and persistent viral infections. Journal of Virology, 66(3), 1442–1448.

Chao, Y.-C., Lee, S.-T., Chang, M.-C., Chen, H.-H., Chen, S.-S., Wu, T.-Y., Liu, F.-H., Hsu, E.-L., & Hou, R. F. (1998). A 2.9-Kilobase Noncoding Nuclear RNA Functions in the Establishment of Persistent Hz-1 Viral Infection. Journal of Virology, 72(3), 2233–2245.

Cheng, C.-H., Liu, S.-M., Chow, T.-Y., Hsiao, Y.-Y., Wang, D.-P., Huang, J.-J., & Chen, H.-H. (2002). Analysis of the Complete Genome Sequence of the Hz-1 Virus Suggests that It Is Related to Members of the Baculoviridae. Journal of Virology, 76(18), 9024–9034. https://doi.org/10.1128/JVI.76.18.9024-9034.2002

Chilcutt, C. F. (2006). Cannibalism of *Helicoverpa zea* (Lepidoptera: Noctuidae) from *Bacillus thuringiensis* (Bt) Transgenic Corn Versus Non-Bt Corn. Journal of Economic Entomology, 99(3), 728–732. https://doi.org/10.1093/jee/99.3.728

Choi, M.-Y., Vander Meer, R. K., Coy, M., & Scharf, M. E. (2012). Phenotypic impacts of PBAN RNA interference in an ant, *Solenopsis invicta*, and a moth, *Helicoverpa zea*. Journal of Insect Physiology, 58(8), 1159–1165. https://doi.org/10.1016/j.jinsphys.2012.06.005

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly, 6(2), 80–92. https://doi.org/10.4161/fly.19695

Cordeiro, E. M. G., Pantoja-Gomez, L. M., de Paiva, J. B., Nascimento, A. R. B., Omoto, C., Michel, A. P., & Correa, A. S. (2020). Hybridization and introgression between *Helicoverpa armigera* and *H. zea*: An adaptational bridge. BMC Evolutionary Biology, 20(1), 61. https://doi.org/10.1186/s12862-020-01621-8

Dader, B., Then, C., Berthelot, E., Ducousso, M., Ng, J.C.K. and Drucker, M. (2017), Insect transmission of plant viruses: Multilayered interactions optimize viral propagation. Insect Science, 24: 929-946. https://doi.org/10.1111/1744-7917.12470

Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: Multiple Alignment of Conserved Genomic Sequence with Rearrangements. Genome Research, 14(7), 1394–1403. https://doi.org/10.1101/gr.2289704

Deng, Z.-L., Dhingra, A., Fritz, A., Götting, J., Münch, P. C., Steinbrück, L., Schulz, T. F., Ganzenmüller, T., & McHardy, A. C. (2021). Evaluating assembly and variant calling software for strain-resolved analysis of large DNA viruses. Briefings in Bioinformatics, 22(3), bbaa123. https://doi.org/10.1093/bib/bbaa123

Dhillon, M. K., & Sharma, H. C. (2013). Comparative studies on the effects of Bt-transgenic and non-transgenic cotton on arthropod diversity, seedcotton yield and bollworms control. Journal of Environmental Biology, 34, 67–73.

Diffenbaugh, N. S., Krupke, C. H., White, M. A., & Alexander, C. E. (2008). Global warming presents new challenges for maize pest management. Environmental Research Letters, 3(4), 044007. https://doi.org/10.1088/1748-9326/3/4/044007

Ditman, L. P., Weiland, G. S., & Guill, J. H., Jr. (1940). The Metabolism in the Corn Earworm. Journal of Economic Entomology, 33(2), 282–295. https://doi.org/10.1093/jee/33.2.282

Drake, V. A. (1984). The vertical distribution of macro-insects migrating in the nocturnal boundary layer: A radar study. Boundary-Layer Meteorology, 28(3), 353–374. https://doi.org/10.1007/BF00121314

Drake, V. A., & Farrow, R. A. (1988). The Influence of Atmospheric Structure and Motions on Insect Migration. Annual Review of Entomology, 33(1), 183–210. https://doi.org/10.1146/annurev.en.33.010188.001151

Drezen, J.-M., Herniou, E. A., & Bézier, A. (2012). Chapter 2—Evolutionary Progenitors of Bracoviruses. In N. E. Beckage & J.-M. Drezen (Eds.), Parasitoid Viruses (pp. 15–31). Academic Press. https://doi.org/10.1016/B978-0-12-384858-1.00002-3

Dubchak, I., Poliakov, A., Kislyuk, A., & Brudno, M. (2009). Multiple whole-genome alignments without a reference organism. Genome Research, 19(4), 682–689. https://doi.org/10.1101/gr.081778.108

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Elias, C. B., Jardin, B., & Kamen, A. (2007). Recombinant Protein Production in Large-Scale Agitated Bioreactors Using the Baculovirus Expression Vector System. In D. W. Murhammer (Ed.), Baculovirus and Insect Cell Expression Protocols (pp. 225–245). Humana Press. https://doi.org/10.1007/978-1-59745-457-5_11

Eliseev, A., Gibson, K. M., Avdeyev, P., Novik, D., Bendall, M. L., Pérez-Losada, M., Alexeev, N., & Crandall, K. A. (2020). Evaluation of haplotype callers for next-generation sequencing of viruses. Infection, Genetics and Evolution, 82, 104277. https://doi.org/10.1016/j.meegid.2020.104277

Engelhard, E. K., Kam-Morgan, L. N. W., Washburn, J. O., & Volkman, L. E. (1994). The Insect Tracheal System: A Conduit for the Systemic Spread of *Autographa californica* M Nuclear Polyhedrosis Virus. Proceedings of the National Academy of Sciences of the United States of America, 91(8), 3224–3227.

Francki RI, Fauquet CM, Knudson DL, Brown F(Eds.). (2012). Classification and nomenclature of viruses: fifth report of the international committee on taxonomy of viruses. Virology Division of the International Union of Microbiological Societies (Vol. 2). Springer Science & Business Media, Germany

Fitt, G. P., Zalucki, M. P., & Twine, P. (1989). Temporal and spatial patterns in pheromone-trap catches of *Helicoverpa* spp. (Lepidoptera: Noctuidae) in cotton-growing areas of Australia. Bulletin of Entomological Research, 79(1), 145–161. https://doi.org/10.1017/S0007485300018654

Friesen, P. D. (1997). Regulation of Baculovirus Early Gene Expression. In L. K. Miller (Ed.), The Baculoviruses (pp. 141–170). Springer US. https://doi.org/10.1007/978-1-4899-1834-5_6

Gasque, S. N., van Oers, M. M., & Ros, V. I. (2019). Where the baculoviruses lead, the caterpillars follow: Baculovirus-induced alterations in caterpillar behaviour. Current Opinion in Insect Science, 33, 30–36. https://doi.org/10.1016/j.cois.2019.02.008

Gilbert, L. I., A. Granger, N., & Roe, R. M. (2000). The juvenile hormones: Historical facts and speculations on future research directions. Insect Biochemistry and Molecular Biology, 30(8), 617–644. https://doi.org/10.1016/S0965-1748(00)00034-5

Gilligan, T. M., Tembrock, L. R., Farris, R. E., Barr, N. B., Straten, M. J. van der, Vossenberg, B. T. L. H. van de, & Metz-Verschure, E. (2015). A Multiplex Real-Time PCR Assay to Diagnose and Separate *Helicoverpa armigera* and *H. zea* (Lepidoptera: Noctuidae) in the New World. PLOS ONE, 10(11), e0142912. https://doi.org/10.1371/journal.pone.0142912

Goldberg, B., Sichtig, H., Geyer, C., Ledeboer, N., & Weinstock, G. M. (2015). Making the Leap from Research Laboratory to Clinic: Challenges and Opportunities for Next-Generation Sequencing in Infectious Disease Diagnostics. MBio, 6(6). https://doi.org/10.1128/mBio.01888-15

Gonçalves, R. M., Mastrangelo, T., Rodrigues, J. C. V., Paulo, D. F., Omoto, C., Corrêa, A. S., & de Azeredo-Espin, A. M. L. (2019). Invasion origin, rapid population expansion, and the lack of genetic structure of cotton bollworm (*Helicoverpa armigera*) in the Americas. Ecology and Evolution, 9(13), 7378–7401. https://doi.org/10.1002/ece3.5123

Granados, R. R., Nguyen, T., & Cato, B. (1978). An insect cell line persistently infected with a baculovirus-like particle. Intervirology, 10(5), 309–317. https://doi.org/10.1159/000148993

Grzywacz, D. (2017). Basic and Applied Research. In Microbial Control of Insect and Mite Pests (pp. 27–46). Elsevier. https://doi.org/10.1016/B978-0-12-803527-6.00003-2

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology, 59(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Hamm, J. J., Carpenter, J. E., & Styer, E. L. (1996). Oviposition Day Effect on Incidence of Agonadal Progeny of *Helicoverpa zea* (Lepidoptera: Noctuidae) Infected with a Virus. Annals of the Entomological Society of America, 89(2), 266–275. https://doi.org/10.1093/aesa/89.2.266

Harding, J. A. (1976). Heliothis spp.: 1 Seasonal Occurrence, Hosts and Host Importance in the Lower Rio Grande Valley 2. Environmental Entomology, 5(4), 666–668. https://doi.org/10.1093/ee/5.4.666

Hardwick, D. (1968). A brief review of the principles of light trap design with a description of an efficient trap for collecting noctuid moths. J. Lepid. Soc, 22(2), 65–75.

Hu, Z., Luijckx, T., van Dinten, L. C., van Oers, M. M., Hajós, J. P., Bianchi, F. J., van Lent, J. W., Zuidema, D., & Vlak, J. M. (1999). Specificity of polyhedrin in the generation of baculovirus occlusion bodies. Journal of General Virology, 80(4), 1045–1053. https://doi.org/10.1099/0022-1317-80-4-1045

Huang, Y.-S., Hedberg, M., & Kawanishi, C. Y. (1982). Characterization of the DNA of a Nonoccluded Baculovirus, Hz-1V. Journal of Virology, 43(1), 174–181.

Huger, A. (1971). Studies on pathological changes of the mid-gut of *Oryctes rhinoceros* infected with *Rhabdionvirus oryctes*. South Pacific Commission, UNDP (SF)/SPC Rhinoceros Beetle Project, Report June 1970—May 1971, 201.

Huger, A. M. (1966). A virus disease of the Indian rhinoceros beetle, *Oryctes rhinoceros* (linnaeus), caused by a new type of insect virus, Rhabdionvirus oryctes gen. N., sp. N. Journal of Invertebrate Pathology, 8(1), 38–51. https://doi.org/10.1016/0022-2011(66)90101-7

Huger, A. M. (2005). The Oryctes virus: Its detection, identification, and implementation in biological control of the coconut palm rhinoceros beetle, *Oryctes rhinoceros* (Coleoptera: Scarabaeidae). Journal of Invertebrate Pathology, 89(1), 78–84. https://doi.org/10.1016/j.jip.2005.02.010

Huger, A. M., & Krieg, A. (1991). Baculoviridae. Nonoccluded Baculoviruses. In Atlas of Invertebrate Viruses. CRC Press.

Ignoffo, C. M., Shapiro, M., & Hink, W. F. (1971). Replication and serial passage of infectious Heliothis nucleopolyhedrosis virus in an established line of *Heliothis zea* cells. Journal of Invertebrate Pathology, 18(1), 131–134. https://doi.org/10.1016/0022-2011(91)90021-H

Iwashita, H., Higa, Y., Futami, K., Lutiali, P. A., Njenga, S. M., Nabeshima, T., & Minakawa, N. (2018). Mosquito arbovirus survey in selected areas of Kenya: Detection of insect-specific virus. Tropical Medicine and Health, 46(1), 19. https://doi.org/10.1186/s41182-018-0095-8

Jehle, J. A. (2010). Nudiviruses. Caister Academic Press: Norwich, UK.

Jiang, L., Goldsmith, M. R., & Xia, Q. (2021). Advances in the Arms Race Between Silkworm and Baculovirus. Frontiers in Immunology, 12, 30. https://doi.org/10.3389/fimmu.2021.628151

Johnson, M. W., Stinner, R. E., & Rabb, R. L. (1975). Ovipositional Response of *Heliothis zea* (Boddie) to Its Major Hosts in North Carolina. Environmental Entomology, 4(2), 291–297. https://doi.org/10.1093/ee/4.2.291

Jones, C. M., Parry, H., Tay, W. T., Reynolds, D. R., & Chapman, J. W. (2019). Movement Ecology of Pest Helicoverpa: Implications for Ongoing Spread. Annual Review of Entomology, 64(1), 277–295. https://doi.org/10.1146/annurev-ento-011118-111959

Jurenka, R. A., Jacquin, E., & Roelofs, W. L. (1991). Control of the pheromone biosynthetic pathway in *Helicoverpa zea* by the pheromone biosynthesis activating neuropeptide. Archives of Insect Biochemistry and Physiology, 17(2–3), 81–91. https://doi.org/10.1002/arch.940170203

Jurenka, R., & Rafaeli, A. (2011). Regulatory Role of PBAN in Sex Pheromone Biosynthesis of Heliothine Moths. Frontiers in Endocrinology, 2, 46. https://doi.org/10.3389/fendo.2011.00046

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kelly, D. C., Lescott, T., Ayres, M. D., Carey, D., Coutts, A., & Harrap, K. A. (1981). Induction of a nonoccluded baculovirus persistently infecting *Heliothis zea* cells by *Heliothis armigera* and *Trichoplusia ni* nuclear polyhedrosis viruses. Virology, 112(1), 174–189. https://doi.org/10.1016/0042-6822(81)90623-1

Khalil, S. M. S., Anspaugh, D. D., & Michael Roe, R. (2006). Role of juvenile hormone esterase and epoxide hydrolase in reproduction of the cotton bollworm, *Helicoverpa zea*. Journal of Insect Physiology, 52(7), 669–678. https://doi.org/10.1016/j.jinsphys.2006.03.004

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. Nature Methods, 12(4), 357–360. https://doi.org/10.1038/nmeth.3317

Kim, K. S., & Kitajima, E. W. (1984). Nonoccluded baculovirus- and filamentous virus-like particles in the spotted cucumber beetle, *Diabrotica undecimpunctata* (Coleoptera: Chrysomelidae). Journal of Invertebrate Pathology, 43(2), 234–241. https://doi.org/10.1016/0022-2011(84)90142-3

Kingan, T. G., Bodnar, W. M., Raina, A. K., Shabanowitz, J., & Hunt, D. F. (1995). The loss of female sex pheromone after mating in the corn earworm moth *Helicoverpa zea*: Identification of a male pheromonostatic peptide. Proceedings of the National Academy of Sciences, 92(11), 5082–5086. https://doi.org/10.1073/pnas.92.11.5082

Kumar Das, J., Tradigo, G., Veltri, P., H Guzzi, P., & Roy, S. (2021). Data science in unveiling COVID-19 pathogenesis and diagnosis: Evolutionary origin to drug repurposing. Briefings in Bioinformatics, bbaa420. https://doi.org/10.1093/bib/bbaa420

Lacey, L. A., Frutos, R., Kaya, H. K., & Vail, P. (2001). Insect Pathogens as Biological Control Agents: Do They Have a Future? Biological Control, 21(3), 230–248. https://doi.org/10.1006/bcon.2001.0938

Larsson, R. (1984). Baculovirus-like particles in the midgut epithelium of the phantom midge, *Chaoborus crystallinus* (Diptera, Chaoboridae). Journal of Invertebrate Pathology, 44(2), 178–186. https://doi.org/10.1016/0022-2011(84)90010-7

Lecuit, M., & Eloit, M. (2015). The potential of whole genome NGS for infectious disease diagnosis. Expert Review of Molecular Diagnostics, 15(12), 1517–1519. https://doi.org/10.1586/14737159.2015.1111140

Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics, 31(10), 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England), 25(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, Y., Farnsworth, C. A., Coppin, C. W., Teese, M. G., Liu, J.-W., Scott, C., Zhang, X., Russell, R. J., & Oakeshott, J. G. (2013). Organophosphate and Pyrethroid Hydrolase Activities of Mutant Esterases from the Cotton Bollworm Helicoverpa armigera. PLoS ONE, 8(10), e77685. https://doi.org/10.1371/journal.pone.0077685

Lin, C.-L., Lee, J.-C., Chen, S.-S., Alan Wood, H., Li, M.-L., Li, C.-F., & Chao, Y.-C. (1999). Persistent Hz-1 Virus Infection in Insect Cells: Evidence for Insertion of Viral DNA into Host Chromosomes and Viral Infection in a Latent Status. Journal of Virology, 73(1), 128–139. https://doi.org/10.1128/JVI.73.1.128-139.1999

Lingren, P. D., Raulston, J. R., Popham, T. W., Wolf, W. W., Lingren, P. S., & Esquivel, J. F. (1995). Flight Behavior of Corn Earworm (Lepidoptera: Noctuidae) Moths Under Low Wind Speed Conditions. Environmental Entomology, 24(4), 851–860. https://doi.org/10.1093/ee/24.4.851

Lipkin, W. I. (2013). The changing face of pathogen discovery and surveillance. Nature Reviews Microbiology, 11(2), 133–141. https://doi.org/10.1038/nrmicro2949

Little, N. S., Luttrell, R. G., Allen, K. C., Perera, O. P., & Parys, K. A. (2017). Effectiveness of Microbial and Chemical Insecticides for Supplemental Control of Bollworm on Bt and Non-Bt Cottons. Journal of Economic Entomology, 110(3), 1039–1051. https://doi.org/10.1093/jee/tow323

López, M. G., Diez, M., Alfonso, V., & Taboga, O. (2018). Biotechnological applications of occlusion bodies of Baculoviruses. Applied Microbiology and Biotechnology, 102(16), 6765–6774. https://doi.org/10.1007/s00253-018-9130-2

Louis, F., Bezier, A., Periquet, G., Ferras, C., Drezen, J.-M., & Dupuy, C. (2013). The Bracovirus Genome of the Parasitoid Wasp *Cotesia congregata* Is Amplified within 13 Replication Units, Including Sequences Not Packaged in the Particles. Journal of Virology, 87(17), 9649–9660. https://doi.org/10.1128/JVI.00886-13

Lundberg, K. S., Shoemaker, D. D., Adams, M. W. W., Short, J. M., Sorge, J. A., & Mathur, E. J. (1991). High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. Gene, 108(1), 1–6. https://doi.org/10.1016/0378-1119(91)90480-Y

Lupiani, B., Raina, A. K., & Huber, C. (1999). Development and Use of a PCR Assay for Detection of the Reproductive Virus in Wild Populations of *Helicoverpa zea* (Lepidoptera: Noctuidae). Journal of Invertebrate Pathology, 73(1), 107–112. https://doi.org/10.1006/jipa.1998.4812

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. Journal, 17(1), 10–12. https://doi.org/10.14806/ej.17.1.200

Matthews, R. (1982). Classification and nomenclature of viruses. Fourth report of the International Committee on Taxonomy of Viruses. Intervirology. https://doi.org/10.1159/ISBN.978-3-8055-8947-5

McIntosh, A. H., Grasela, J. J., & Ignoffo, C. M. (2007). In vitro host range of the Hz-1 nonoccluded virus in insect cell lines. In Vitro Cellular & Developmental Biology. Animal, 43(5–6), 196–201. https://doi.org/10.1007/s11626-007-9032-6

McIntosh, A. H., Ignoffo, C. M., & Andrews, P. L. (1985). In vitro host range of five baculoviruses in lepidopteran cell lines. Intervirology, 23(3), 150–156.

McNeil, J. N., Cusson, M., Delisle, J., Orchard, I., & Tobe, S. S. (1995). Physiological integration of migration in Lepidoptera. In A. G. Gatehouse & V. A. Drake (Eds.), Insect Migration: Tracking Resources through Space and Time (pp. 279–302). Cambridge University Press. https://doi.org/10.1017/CBO9780511470875.014

Miller, L. K., & Ball, L. A. (1998). The Insect Viruses. Springer US. http://public.eblib.com/choice/publicfullrecord.aspx?p=3081418

Moore, V. M., & Tracy, W. F. (2021). Survey of organic sweet corn growers identifies corn earworm prevalence, management and opportunities for plant breeding. Renewable Agriculture and Food Systems, 36(2), 126–129. https://doi.org/10.1017/S1742170520000204

Muller, H., Chebbi, M. A., Bouzar, C., Périquet, G., Fortuna, T., Calatayud, P.-A., Le Ru, B., Obonyo, J., Kaiser, L., Drezen, J.-M., Huguet, E., & Gilbert, C. (2021). Genome-wide patterns of bracovirus chromosomal integration into multiple host tissues during parasitism. Journal of Virology. https://doi.org/10.1128/JVI.00684-21

Öhlund, P., Lundén, H., & Blomström, A.-L. (2019). Insect-specific virus evolution and potential effects on vector competence. Virus Genes, 55(2), 127–137. https://doi.org/10.1007/s11262-018-01629-9

Okonechnikov, K., Golosova, O., & Fursov, M. (2012). Unipro UGENE: A unified bioinformatics toolkit. Bioinformatics, 28(8), 1166–1167. https://doi.org/10.1093/bioinformatics/bts091

Parize, P., Muth, E., Richaud, C., Gratigny, M., Pilmis, B., Lamamy, A., Mainardi, J.-L., Cheval, J., de Visser, L., Jagorel, F., Ben Yahia, L., Bamba, G., Dubois, M., Join-Lambert, O., Leruez-Ville, M., Nassif, X., Lefort, A., Lanternier, F., Suarez, F., Eloit, M. (2017). Untargeted next-generation sequencing-based first-line diagnosis of infection in immunocompromised adults: A multicentre, blinded, prospective study. Clinical Microbiology and Infection, 23(8), 574.e1-574.e6. https://doi.org/10.1016/j.cmi.2017.02.006

Pearce, S. L., Clarke, D. F., East, P. D., Elfekih, S., Gordon, K. H. J., Jermiin, L. S., McGaughran, A., Oakeshott, J. G., Papanikolaou, A., Perera, O. P., Rane, R. V., Richards, S., Tay, W. T., Walsh, T. K., Anderson, A., Anderson, C. J., Asgari, S., Board, P. G., Bretschneider, A., Wu, Y. D. (2017). Genomic innovations, transcriptional plasticity and gene loss underlying the evolution and divergence of two highly polyphagous and invasive Helicoverpa pest species. BMC Biology, 15(1), 63. https://doi.org/10.1186/s12915-017-0402-6

Perera, O. P., Allen, K. C., Jain, D., Purcell, M., Little, N. S., & Luttrell, R. G. (2015). Rapid Identification of *Helicoverpa armigera* and *Helicoverpa zea* (Lepidoptera: Noctuidae) Using Ribosomal RNA Internal Transcribed Spacer 1. Journal of Insect Science, 15(1), 155. https://doi.org/10.1093/jisesa/iev137

Pogue, M. G. (2004). A New Synonym of Helicoverpa zea (Boddie) and Differentiation of Adult Males of *H. zea* and *H. armigera* (Hübner) (Lepidoptera: Noctuidae: Heliothinae). Annals of the Entomological Society of America, 97(6), 1222–1226. https://doi.org/10.1603/0013-8746(2004)097[1222:ANSOHZ]2.0.CO;2

Popham, H. J. R., Nusawardani, T., & Bonning, B. C. (2016). Introduction to the Use of Baculoviruses as Biological Insecticides. In D. W. Murhammer (Ed.), Baculovirus and Insect Cell Expression Protocols (pp. 383–392). Springer. https://doi.org/10.1007/978-1-4939-3043-2_19

Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., & Korobeynikov, A. (2020). Using SPAdes De Novo Assembler. Current Protocols in Bioinformatics, 70(1), e102. https://doi.org/10.1002/cpbi.102

R Core Team. (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org/

Rafaeli, A., & Bober, R. (2005). The effect of the juvenile hormone analog, fenoxycarb on the PBAN-receptor and pheromone production in adults of the moth *Helicoverpa armigera*: An "aging" hormone in adult females? Journal of Insect Physiology, 51(4), 401–410. https://doi.org/10.1016/j.jinsphys.2005.01.004

Raina, A. K., & Adams, J. R. (1995). Gonad-specific virus of corn earworm. Nature, 374(6525), 770–770. https://doi.org/10.1038/374770a0

Raina, A. K., Adams, J. R., Lupiani, B., Lynn, D. E., Kim, W., Burand, J. P., & Dougherty, E. M. (2000). Further Characterization of the Gonad-Specific Virus of Corn Earworm, Helicoverpa zea. Journal of Invertebrate Pathology, 76(1), 6–12. https://doi.org/10.1006/jipa.2000.4942

Rallis, C. P., & Burand, J. P. (2002a). Pathology and ultrastructure of Hz-2V infection in the agonadal female corn earworm, *Helicoverpa zea*. Journal of Invertebrate Pathology, 81(1), 33–44. https://doi.org/10.1016/S0022-2011(02)00113-1

Rallis, C. P., & Burand, J. P. (2002b). Pathology and ultrastructure of the insect virus, Hz-2V, infecting agonadal male corn earworms, *Helicoverpa zea*. Journal of Invertebrate Pathology, 80(2), 81–89. https://doi.org/10.1016/S0022-2011(02)00102-7

Ralston, A. L., Huang, Y.-S., & Kawanishi, C. Y. (1981). Cell culture studies with the IMC-Hz-1 nonoccluded virus. Virology, 115(1), 33–44. https://doi.org/10.1016/0042-6822(81)90086-6

Ramaswamy, S. B., Shu, S., Park, Y. I., & Zeng, F. (1997). Dynamics of juvenile hormone-mediated gonadotropism in the lepidoptera. Archives of Insect Biochemistry and Physiology, 35(4), 539–558. https://doi.org/10.1002/(SICI)1520-6327(1997)35:4<539::AID-ARCH12>3.0.CO;2-B

Reisig, D. D., Huseth, A. S., Bacheler, J. S., Aghaee, M.-A., Braswell, L., Burrack, H. J., Flanders, K., Greene, J. K., Herbert, D. A., Jacobson, A., Paula-Moraes, S. V., Roberts, P., & Taylor, S. V. (2018). Long-Term Empirical and Observational Evidence of Practical *Helicoverpa zea* Resistance to Cotton with Pyramided Bt Toxins. Journal of Economic Entomology, 111(4), 1824–1833. https://doi.org/10.1093/jee/toy106

Rogers, C. D. G., & Burgoyne, L. A. (2000). Reverse transcription of an RNA genome from databasing paper (FTA®). Biotechnology and Applied Biochemistry, 31(3), 219–224. https://doi.org/10.1042/BA19990113

Ronquist, F., Teslenko, M., Mark, P. van der, Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Systematic Biology, 61(3), 539–542. https://doi.org/10.1093/sysbio/sys029

Sajjan, D. B., & Hinchigeri, S. B. (2016). Structural Organization of Baculovirus Occlusion Bodies and Protective Role of Multilayered Polyhedron Envelope Protein. Food and Environmental Virology, 8(1), 86–100. https://doi.org/10.1007/s12560-016-9227-7

Sambrook, J., & Russell, D. W. (2006). Standard Ethanol Precipitation of DNA in Microcentrifuge Tubes. Cold Spring Harbor Protocols, 2006(1), pdb.prot4456. https://doi.org/10.1101/pdb.prot4456

Sandstrom, M. A., Changnon, D., & Flood, B. R. (2007). Improving our Understanding of *Helicoverpa zea* Migration in the Midwest: Assessment of Source Populations. Plant Health Progress, 8(1), 63. https://doi.org/10.1094/PHP-2007-0719-08-RV

Shapiro-ilan, D. I., Gardner, W. A., Fuxa, J. R., Wood, B. W., Nguyen, K. B., Adams, B. J., Humber, R. A., & Hall, M. J. (2003). Survey of Entomopathogenic Nematodes and Fungi Endemic to Pecan Orchards of the Southeastern United States and Their Virulence to the Pecan Weevil (Coleoptera: Curculionidae). Environmental Entomology, 32(1), 187–195. https://doi.org/10.1603/0046-225X-32.1.187

Shim, H. J., Choi, J. Y., Wang, Y., Tao, X. Y., Liu, Q., Roh, J. Y., Kim, J. S., Kim, W. J., Woo, S. D., Jin, B. R., & Je, Y. H. (2013). NeuroBactrus, a Novel, Highly Effective, and Environmentally Friendly Recombinant Baculovirus Insecticide. Applied and Environmental Microbiology, 79(1), 141–149. https://doi.org/10.1128/AEM.02781-12

Slack, J., & Arif, B. M. (2006). The Baculoviruses Occlusion-Derived Virus: Virion Structure and Function. In Advances in Virus Research (Vol. 69, pp. 99–165). Academic Press. https://doi.org/10.1016/S0065-3527(06)69003-9

Smith, K. M. (1967). Insect Virology. Elsevier Science.

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evolution, 4(vey016). https://doi.org/10.1093/ve/vey016

Sutton, T. D. S., Clooney, A. G., Ryan, F. J., Ross, R. P., & Hill, C. (2019). Choice of assembly software has a critical impact on virome characterisation. Microbiome, 7(1), 12. https://doi.org/10.1186/s40168-019-0626-5

Tabashnik, B. E., Liesner, L. R., Ellsworth, P. C., Unnithan, G. C., Fabrick, J. A., Naranjo, S. E., Li, X., Dennehy, T. J., Antilla, L., Staten, R. T., & Carrière, Y. (2021). Transgenic cotton and sterile insect releases synergize eradication of pink bollworm a century after it invaded the United States. Proceedings of the National Academy of Sciences, 118(1). https://doi.org/10.1073/pnas.2019115118

Thézé, J., Bézier, A., Periquet, G., Drezen, J.-M., & Herniou, E. A. (2011). Paleozoic origin of insect large dsDNA viruses. Proceedings of the National Academy of Sciences, 108(38), 15931–15935. https://doi.org/10.1073/pnas.1105580108

Treuhaft, M. W., & Beem, M. O. (1982). Defective interfering particles of respiratory syncytial virus. Infection and Immunity, 37(2), 439–444.

Vago, C. (1963). A new type of insect virus. J. Insect Pathol., 5, 275–276.

Vignuzzi, M., & López, C. B. (2019). Defective viral genomes are key drivers of the virus–host interaction. Nature Microbiology, 4(7), 1075–1087. https://doi.org/10.1038/s41564-019-0465-y

Volkman, L. E. (1997). Nucleopolyhedrovirus Interactions with Their Insect Hosts. In K. Maramorosch, F. A. Murphy, & A. J. Shatkin (Eds.), Advances in Virus Research (Vol. 48, pp. 313–348). Academic Press. https://doi.org/10.1016/S0065-3527(08)60291-2

Wang, Y., Burand, J. P., & Jehle, J. A. (2007). Nudivirus genomics: Diversity and classification. Virologica Sinica, 22(2), 128–13. https://doi.org/10.1007/s12250-007-0014-3

Wang, Y., & Jehle, J. A. (2009). Nudiviruses and other large, double-stranded circular DNA viruses of invertebrates: New insights on an old topic. Journal of Invertebrate Pathology, 101(3), 187–193. https://doi.org/10.1016/j.jip.2009.03.013

Wang, Y., Bininda-Emonds R.P., & Jehle, A. (2012). Nudivirus Genomics and Phylogeny. In M. Garcia (Ed.), Viral Genomes—Molecular Structure, Diversity, Gene Expression Mechanisms and Host-Virus Interactions. InTech. https://doi.org/10.5772/27793

Wang, Y., van Oers, M. M., Crawford, A. M., Vlak, J. M., & Jehle, J. A. (2007). Genomic analysis of *Oryctes rhinoceros* virus reveals genetic relatedness to Heliothis zea virus 1. Archives of Virology, 152(3), 519–531. https://doi.org/10.1007/s00705-006-0872-2

Ward, R. D. C. (1916). The Prevailing Winds of the United States. Annals of the Association of American Geographers, 6, 99–119. https://doi.org/10.2307/2569453

Westbrook, J. K., Eyster, R. S., Wolf, W. W., Lingren, P. D., & Raulston, J. R. (1995). Migration pathways of corn earworm (Lepidoptera: Noctuidae) indicated by tetroon trajectories. Agricultural and Forest Meteorology, 73(1), 67–87. https://doi.org/10.1016/0168-1923(94)02171-F

Westbrook, J. K., & Jr, J. D. L. (2010). Long-Distance Migration in Helicoverpa zea1: What We Know and Need to Know. Southwestern Entomologist, 35(3), 355–360. https://doi.org/10.3958/059.035.0315

Widstrom, N. W., Lillehoj, E. B., Sparks, A. N., & Kwolek, W. F. (1976). Corn Earworm Damage and Aflatoxin B1 on Corn Ears Protected with Insecticide. Journal of Economic Entomology, 69(5), 677–679. https://doi.org/10.1093/jee/69.5.677

Wongteerasupaya, C., Vickers, J., Sriurairatana, S., Nash, G., Akarajamorn, A., Boonsaeng, V., Panyim, S., Tassanakajon, A., Withyachumnarnkul, B., & Flegel, T. (1995). A non-occluded, systemic baculovirus that occurs in cells of ectodermal and mesodermal origin and causes high mortality in the black tiger prawn *Penaeus monodon*. https://doi.org/10.3354/DAO021069

Wu, P.-C., Lin, Y.-H., Wu, T.-C., Lee, S.-T., Wu, C.-P., Chang, Y., & Wu, Y.-L. (2018). MicroRNAs derived from the insect virus Hz NV-1 promote lytic infection by suppressing histone methylation. Scientific Reports, 8(1), 17817. https://doi.org/10.1038/s41598-018-35782-w

Wu, Y.-L., Liu, C. Y. Y., Wu, C. P., Wang, C.-H., Lee, S.-T., & Chao, Y.-C. (2008). Cooperation of ie1 and p35 genes in the activation of baculovirus AcMNPV and HzNV-1 promoters. Virus Research, 135(2), 247–254. https://doi.org/10.1016/j.virusres.2008.04.001

Wu, Y.-L., Wu, C. P., Lee, S.-T., Tang, H., Chang, C.-H., Chen, H.-H., & Chao, Y.-C. (2010). The Early Gene hhi1 Reactivates *Heliothis zea* Nudivirus 1 in Latently Infected Cells. Journal of Virology, 84(2), 1057–1065. https://doi.org/10.1128/JVI.01548-09

Wu, Y.-L., Wu, C. P., Liu, C. Y. Y., Lee, S.-T., Lee, H.-P., & Chao, Y.-C. (2011). Heliothis zea Nudivirus 1 Gene hhi1 Induces Apoptosis Which Is Blocked by the Hz-iap2 Gene and a Noncoding Gene, pag1. Journal of Virology, 85(14), 6856–6866. https://doi.org/10.1128/JVI.01843-10

Yuan, C., Xing, L., Wang, M., Hu, Z., & Zou, Z. (2021). Microbiota modulates gut immunity and promotes baculovirus infection in *Helicoverpa armigera*. Insect Science. https://doi.org/10.1111/1744-7917.12894

VITA

Emrah Ozel

Education

- Hacettepe University, Dept. of Biology, 2001 – 2007, Bachelor of Science.
- Hacettepe University, Institute of Natural Sciences, 2007 – 2010, Master of Science.
- Texas State University, Aquatic Resources Master's Program.
- University of Kentucky, Entomology PhD Program,

Work Experience

- Cave Researcher

  2009, Sustainable Development and Protection of Natural Resources and Biodiversity of Yildiz Mountains Project. (UNDP)
- Coordinator

  2010, Determination of Invertebrate Fauna and Surveying Zoogeographic Variations of Balıkesir – Havran Region Caves Project (TUBITAK).
- Teaching Assistant

  2011 – 2012, Department of Biology (Texas State University - San Marcos)
- Research Assistant
- 2012 – 2015, Department of Biology (Texas State University - San Marcos)
- Laboratory Technician

  2016 – 2018, Entomology Department, University of Kentucky (Supervisor: Dr. Bruce Webb).
- Research Assistant

  2018 – 2021, Entomology Department, University of Kentucky (Supervisor: Dr. Bruce Webb).

Publications

1. Brown, D. J., D. B. Preston, E. Ozel, and M. R. J. Forstner. 2013. Wildfire impacts on fire ant captures around forest ponds in the Lost Pines of Texas. Journal of Fish and Wildlife Management 4(1):129-133.

2. Brown, D. J., W. H. Nowlin, E. Ozel, I. Mali, D. Episcopo, M. C. Jones, and M. R. J. Forstner. 2014. Comparison of short term low, moderate, and high severity fire impacts to aquatic and terrestrial habitat components of a southern USA mixed pine forest. Forest Ecology and Management 312(2014):179-192.
3. [In prep.] Heliothine nudivirus Prevalence in the Cotton Belt Region and Surrounding States
4. [In prep.] A Digital Survey of Helicoverpa zea Nudivirus in public sequence databases
5. [In prep.] A novel nudivirus that infects Heliothis virescens populations in the Cotton Belt region

Presentations

1. Donald Brown, Devin Preston, Emrah Ozel, Michael Forstner (2012); The Effect of Canopy Cover, Prescribed Fire, Temperature and Precipitation on Red Imported Fire Ant (Solenopsis invicta, RIFA) Activity in the Lost Pines Ecoregion of Texas. Texas Herpetological Society.
2. Donald Brown, Devin Preston, Emrah Ozel, Michael Forstner (2012); How do Fire Ants (Solenopsis invicta) respond to the changes brought on by last years wildfires in the Lost Pines region of Texas. International Research Conference for Graduate Students, Texas State University.
3. Emrah Ozel, Michael Forstner (2014); Exposure to RIFA may be a higher threat than previously considered: Seasonal habitat shifts of red imported fire ant (Solenopsis invicta) negatively affects Houston toad distribution. (POSTER), 2014 THS Fall Symposium, Texas State University.
4. Emrah Ozel, Bruce Webb (2019); H. zea Nudivirus-2: A Novel Control Agent for Corn Earworm (Helicoverpa zea) Management, ESA Annual Meeting, St. Louis, MO.
5. Emrah Ozel, Bruce Webb (2020); A bioinformatic survey of Helicoverpa zea Nudivirus, ESA Annual Meeting, Virtual.

Reports

1. Michael R.J. Forstner, S. McCracken, and E. Ozel (2014); Minimizing Wildlife-Motorist Interactions, Annual Report for Texas Department of Transportation. Texas State University, San Marcos.